The Dissertation Committee for Helen Louise Johnson
certifies that this is the approved version of the following dissertation

CONSEQUENCES OF HIGH-STAKES TESTING: CRITICAL PERSPECTIVES

OF TEACHERS AND STUDENTS

Committee:

_____
Jay Scribner, Supervisor

_____
Lisa Cary

_____
Terry Clark

_____
Pat Forgione

_____
Norvell Northcutt

CONSEQUENCES OF HIGH-STAKES TESTING: CRITICAL PERSPECTIVES OF TEACHERS AND STUDENTS

by

Helen Louise Johnson, B.A., M.Ed.

Dissertation

Presented to the Faculty of the Graduate School
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of
Doctor Of Education

The University of Texas at Austin

August 2004

DEDICATION

This dissertation is dedicated to memory of my grandparents, Rev. Mark E. and Charity Smith, and of my parents, Walker and Ruby Fleming, who modeled determination and strove for educational excellence. It is also dedicated to my children Andrew, Heather, and Nozomi Johnson, whom I owe much and expect to continue the family tradition of perseverance and higher education.

ACKNOWLEDGEMENTS

I would like to acknowledge the outstanding support I have received from the faculty and staff of the College of Education, Department of Educational Administration, Executive Leadership Program. A special word of thanks goes to Dr. Nolan Estes, for his invaluable expertise, inspiration, and teachings, for which I am eternally grateful. Deep gratitude also goes to Dr. Larry Lewis, Dr. Barbara Freeman, Dr. Yvonne Johnson, and Dr. Charity Smith for their encouragement and support in my pursuit of the doctoral degree.

I am likewise indebted to Dr. Jay Scribner, who encouraged me to pursue this study and who guided me through the dissertation process. And I would like to thank him and the other committee members—Dr. Lisa Cary, Dr Terry Clark, Dr. Pat Forgione, and Dr. Novell Northcutt—for allowing me to make mistakes, challenging my thinking, and providing constructive criticism to improve the work.

This work would not have been accomplished without the cooperation of the Angleton School District. I would like to recognize the outstanding support I received. And I am indebted to all those who shared their experiences, time, and resources.

The personal support I received from my friends, family, and community was incredible. I want to thank my husband, Andrew Johnson, and my children for their love, support, and understanding. I am deeply appreciative of the help I have received from sisters and brothers: Lawrence Fleming, Walker Fleming, Charles Fleming, Georgia State Representative Carolyn Hugley, Ruby Bridgeforth, Dr. Charity Smith, and Everett Fleming. And I was blessed to have

the love and caring of a host of cousins, aunts, and uncles, especially that of my aunt and uncle, Elporter and Artelelure Gamble.

Consequences of High-Stakes Testing: Critical Perspectives of Teachers and Students

Publication No. _____

Helen Louise Johnson, Ed.D.
The University of Texas at Austin, 2004

Supervisor: Jay Scribner

High-stakes testing has been promised by its proponents as the way to make American schools accountable. In theory, high-stakes testing is commendable. In practice, many minority and poor high school students fail to pass high-stakes exams and graduate as abundant research indicates.

Previous studies of high-stakes testing often have tended to overlook or have failed to examine the unintended consequences of high-stakes testing. The purpose of this study is to determine the unintended consequences of high-stakes testing—specifically, side effects of high-stakes testing for economically

disadvantaged and ethnic minority students, from teachers' and students' points of view.

High-stakes testing involves educational research for public schools. Three questions have guided this research: What critical perspectives do teachers (and students) have about the effects of high-stakes testing? How do intended and unintended consequences of high-stakes testing affect instruction? How do intended and unintended consequences of high-stakes testing affect student outcomes?

Methodology for the study involves qualitative research via a case study of an urban central Texas school district, employing simple random sampling to select study informants, and using Interactive Qualitative Analysis (IQA) as a method of data analysis.

The study is intended to contribute to the improvement of professional practices in public schools and to educational research.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I – INTRODUCTION

Assessment has been engraved in the educational landscape of America since the early 1900s. However, the purpose of testing has changed from a private measure of individual student performance to highly public, publicized measures of schools', districts', and states' accountability with rewards and/or punishments for both institutions and individuals. High-stakes tests are assessments from which results are used to make significant educational decisions about schools, teachers, administrators, and students (Amrein & Berliner, 2002). High-stakes tests have been promised to America as the mechanism to foster educational reform—specifically, that of improving student learning—and to hold schools accountable. However, high-stakes testing policies disproportionately affect many African-American, Hispanic, and poor students. Minorities claim that changes in the high-stakes testing policy permit system structural and institutional mechanisms that cause and promote unfair consequences for minority students. This chapter discusses the context of the study, problem being investigated, study methodology, and significance of the study.

Context of the Study

James Baldwin (Tatum, 1999) informs us, "Not everything that is faced can be changed. But nothing can be changed until it is faced" (p.20). Although his remark was not referring to high-stakes assessment, no statement could be more applicable to the high-stakes testing dilemma. In this era of accountability and assessment, there is strong support for public policies that use high-stakes testing to change the behavior of teachers and students in desirable ways.

According to Corbett and Wilson (1991), the consequences of testing can be both intended and unintended. The potential problem with the current increased emphasis on testing is not necessarily the test, per se, but the instances when tests have unintended and potentially negative consequences for individual students, groups of students, or the educational system (*Appropriate use of high-stakes testing in our nation's schools*, 2003). Corbett and Wilson (1991) state:

> Stakes can become high when test results automatically trigger important consequences for students or the school system, and also when educators, students, or the public perceive that significant consequence accompany test results. Thus, a formal trigger of consequences need not be built into the testing program for stakes to be high
>
> . . . The product of this process can be increased public pressure to improve test scores, especially when the perception is that the system is likely to have a negative impact on those choices (p. 27).

Legislators and policymakers sanction the use of high-stakes testing with the intent to use tests in setting high standards for student learning, raising student achievement-levels, ensuring equal educational opportunity, fostering parental involvement in student learning, and increasing public support for schools (*AERA position statement concerning high-stakes testing in pre-K-12 education*, 2002; Heubert & Hauser, 1999). Tests are intended to increase student performance, make schools accountable, and close the achievement gaps between whites and minorities.

"Stakes" or consequences associated with test results have long been a part of American policies. During the twentieth century, scores on high-stakes tests could result in such outcomes as: immigrants gaining entrance to or being rejected from the United States; public school students being placed in gifted programs, vocational tracks, or even in the homes for the mentally inferior. Used in this way, the consequences of standardized testing ensured maintenance of

the status quo along those racial, ethnic, and class lines (Amrein & Berliner, 2002).

Legislators promote high-stakes testing and accountability for all; however, schools that historically perform well on exams are not targets of these policies; poor urban under-performing schools are. In successful communities, high-stakes testing is acceptable because it confirms the expectations of the community, posing little threat, and also has little incentive value. In poorer communities, high-stakes policies are supported because, at the very least, educational standards are raised.

However, if high-stakes testing policies do not promote learning, leading to education in the most profound sense, then the test will not be of any use in successful communities and schools, nor will they improve the schools attended by poor children or ethnic minorities. Additionally, the test will have unintended consequences, such as narrowing curriculum, increasing dropout rate, and contributing to the higher retention rate (Amrein & Berliner, 2002). Thus, unintended consequences of high-stakes testing may be defined as unexpected or negative consequences that impact individual students, groups of students, or the educational system, often causing:

- narrowed curriculum
- increased dropout rate
- higher retention rates
- diminished educational opportunities for students
- decreased support for public schools.

It is important to remember that, without testing, many low-performing schools would remain invisible. Determining whether high-stakes testing of

students produces better outcomes requires that its potential benefits be weighed against its potential unintended or negative consequences.

High-stakes testing has become a key component of educational reform. According to American Educational Research Association (AERA), pressure to raise test scores can force states—and district-level officials—to make decisions that may run contrary to what is best for students (*What the AERA says about high-stakes testing, 2002*), causing negative or unintended consequences. AERA contends that decisions which affect individual students' life chances or educational opportunities should not be made on the basis of test scores alone (*What the AERA says about high-stakes testing, 2002*).

High-stakes testing is becoming an important educational hurdle for high school students, specifically, poor and minority students. Passing or failing of high-stakes exams often determines students' future access to educational and/or job opportunities. States often sell high-stakes testing as a tool to assure accountability. The logic of having an exit examination linked to high school graduation makes sense in a perfect world in which equal educational opportunities exist. However, few students live in perfect environments.

From an institutional perspective, policymakers and some members of the public view high-stakes testing as a way to improve the quality of schools in a cost-effective and efficient manner. Culturalists and others have described high-stakes testing as legalized child abuse, where perpetrators are neither clearly identified nor held accountable and children are unfairly punished. Regardless of the point of view, high-stakes testing is quickly being recognized as the key for children to access future opportunities (Valencia, Valenzuela, Sloan, & Foley 2001).

The reauthorization of the Elementary and Secondary Education Act in 1994 reformed federal programs to support state efforts to establish challenging standards, develop aligned assessments, and build accountability systems for districts and schools that are based on educational results (*Peer reviewer guidance for evaluating evidence of final assessments under Title I of the Elementary and Secondary Education Act*, 1999). The Act includes explicit requirements to ensure that students served by Title I are given the same opportunity to achieve high standards and are held to the same expectations as all students in the state. In addition, the public demand for high standards and improved student performance has caused many states in the union to create and administer tests that measure student performance over time. For growing numbers of students in the United States, a plain paper test booklet has become a powerful gatekeeper of their future. According to the Center on Education Policy, currently 18 states, enrolling half of all public school students, require their students to pass exit examinations before they can graduate from high school and receive a regular diploma (Chudowsky, Kober, Gayler, & Hamilton, 2002). It is projected that, within six years, at least 24 states will have mandatory exit exams, affecting about seven of every ten public school students and eight of ten minority students (Chudowsky et al., 2002).

From a macro-political viewpoint, the stated objective of the “standards” movement in American public education is to hold all schools, teachers and students to high standards of learning. The movement reflects awareness that student proficiencies in literacy and mathematics largely determine success in school and employment (Murnane & Levy, 1996; Sum, 1999). This movement has caused the expansion of exit exams. State policymakers are often the main

drivers of this effort. Standards have given a more solid foundation to the concept of exit exams by laying out what students should know and be able to do by the time they graduate from high school. Often, state tests are used as yardsticks for measuring student progress and as instruments for holding students and educators accountable for high performance (Chudowsky et al., 2002).

Many states are adopting exit exams in an effort to make high school diplomas “have value,” namely, to validate that the holder has the knowledge and skills needed to do well in a job or in college. State policymakers are adopting these exams to respond to public outcry that the quality of public education has slipped and too many students are graduating with substandard or no academic skills. Advocates say these exams will motivate students to work harder, and help teachers to identify and address student weaknesses. Indeed, some analyses of national test results suggest that those states such as North Carolina and Texas showed large gains in mathematics between 1992 and 1996 (Grissmer & Flanagan, 1998).

Critics argue that such testing does not promote real improvements in student learning. Rather, teachers and principals are motivated to meet standards by “teaching the test” instead of creating an improved learning environment. These crude forms of assessment may reduce opportunities to learn higher-order skills (McNeil & Valenzuela, 2000). Critics also contend that these tests lead to higher dropout rates, place too much weight on a single measure, and do nothing to ensure that students have the opportunity to learn the material before testing. Additionally, they claim that state testing increases the probability that disadvantaged students will drop out of school by forcing students to repeat grades (Haney, 1999, 2000).

The Statement of Problem

Assessment and accountability are now prominent features of the American educational system. High-stakes assessment is a crucial part of accountability. Many students begin high school dreaming of successfully completing state exit exams and graduating. However, passing state high school exit exams is becoming the single most important hurdle for poor, African-American, and Hispanic children. Studies of high-stakes testing often overlook the unintended consequences of these assessments in an effort to support such testing. However, few studies have rigorously examined the consequences of high-stakes testing from teachers and students' points of view. According to Amrein and Berliner, it is now time to debate high-stakes testing policies more thoroughly, and seek to change them if they do not do what was intended and if they have some unintended negative consequences (Amrein & Berliner, 2002). Thus, the unintended consequences of high-stakes testing on high school students, specifically economically disadvantaged and ethnic minority students need to be explored.

The Purpose

The purpose of this study is to determine the consequences of high-stakes testing on high school students, specifically the unintended side effects of high-stakes testing for economically disadvantaged and ethnic minority students, from the critical perspective of teachers and students. It is important to study these intended and unintended consequences for students during large scale assessments for the following reasons: (1) high-stakes testing policy is acclaimed as the legislated instrument to improve teaching and learning—the ultimate goal of education reforms—by making schools more accountable for the performance

of their students; (2) the possibility exists that high-stakes testing policy may be a symbolic response to very real educational problems (Ellwein, Glass, & Smith, 1988); (3) the need exists to examine how policy made at the macro-political (state and federal) levels affects students at the micro-political level; (4) legislated policy responses to educational problems often occur without sufficient study of the efficacy of the policy, its effect, and its appropriateness to identify the problem (Ellwein et al., 1988); and (5) the need exists to examine fairness and equity of high-stakes testing policies for all students. Further, strategies to eliminate and/or reduce the unintended consequences of high-stakes assessment meant to ensure that all students meet and master educational standards given by teachers will be examined.

Research Questions

The research is organized around three related research questions. These questions are:

- What critical perspectives do teachers (and students) have about the effects of high-stakes testing?
- How do intended and unintended consequences of high-stakes testing affect students?

▪ How do intended and unintended consequences of high-stakes testing affect instruction and classroom practices?

Methodology

This is a qualitative research study involving a case study of a central Texas urban school district within a political educational framework with a policy paradigm. It focuses on the effects and consequences of high-stakes testing

policy on students—specifically, unintended consequences for minority students—from critical perspective of teachers and students. Instruments used in the study are state achievement tests, focus groups, partially structured interviews, online questionnaires, and participant observations. Triangulation, verbatim accounts of observation and interviews, and research of daily reflections will be used as data collection validity checks.

Interactive Qualitative Analysis (IQA), a unique innovative research design, was chosen for the study because it allows the researcher to perform a more in-depth assessment of studied phenomena. The circular nature of IQA with recursive (looping) features allows for successive project refinement, transforming the vague outcomes often found in linear qualitative research methods into more and more precise, relevant outcomes. The problem is investigated from the viewpoints of constituencies, power and distance, issues, comparisons, and research questions.

Definitions of General Terms

Criterion Reference Test. *Criterion reference test* is an assessment that measures a student's performance according to specified standards or criteria rather than in comparison to the performance of other test-takers.

Clarity of users. The condition whereby alignment of the standards and the assessment is clear to all members of the school community.

Comprehensiveness. The quality of assessments that reflect the full range of the standards, and that are complete by other measures such as another test or local measures that provide information to educators on the other standards.

Depth. *Depth* refers to standards that are as cognitively demanding as the assessments.

Descriptive research. "Descriptive research involves the collection and analysis of quantitative data in order to develop a precise description of a sample's behavior or personal characteristics" (Gall, Gall, & Borg, 1999).

Education Capital. *Education capital* is the development of thinking skills that will develop learner-products as powerful contributors in the 21st century market place.

Emphasis. A quality of assessments such that they reflect the same degree of importance on the different content standards as reflected in the standards documents.

Fairness/accessibility. *Fairness* refers to ensuring that all students have an equal opportunity to show what they can do, in spite of the fact they have different backgrounds, different and complex patterns of abilities that interact with the assessment process itself, and different opportunities to meet these assessments.

High-stakes Decision. A *high-stakes decision* is a single decision that affects an individual student's life chances or educational opportunities, such as promotion to the next grade or graduation from high school.

High-stakes Test. A test can be considered *high-stakes* if the results of the test produce or cause perceived or real consequences for students, staff, or schools (Madaus, 1988). High-stakes tests are tests used in making decisions about which students will be promoted or retained in grade and which will receive high-school diplomas.

Learning. Learning occurs when the child's behavior is changed in the direction of the desired learner outcomes (Foster, 1985).

Impacts. *Impact* refers to risks and/or consequences of high-stakes tests.

Interpretation and use. The term *interpretation and use* refers to ensuring that users of the assessment data have the support needed to draw the most appropriate interpretations and use the results in the most valid ways.

Reliability. *Reliability* occurs the level of consistency, stability and accuracy of the assessment.

Texas Assessment of Academic Skills *(TAAS).* The TAAS is a criterion-referenced test used in Texas from 1990–2001 to assess academic skills primarily in mathematics, reading and writing (Texas Education Agency, 1993).

Texas Assessment of Basic Skills *(TABS).* The TABS is a criterion-referenced test used in Texas before 1985 to assess basic skills in mathematics, reading and writing (Texas Education Agency, 1993).

Texas Education Assessment of Minimum Skills *(TEAMS).* The TEAMS is a criterion-referenced test used in Texas from 1985–1990 to assess basic skills in mathematics, reading and writing (Texas Education Agency, 1993).

Texas Assessment of Knowledge and Skills *(TAKS).* The TAKS is a criterion-referenced test used in Texas from 2002 to present to assess academic skills primarily in mathematics, reading and writing, science, and social studies (*Texas Assessment of Knowledge and Skills (TAKS) - Update*, 2002).

Survey Research. "Survey research is a form of descriptive research that involves collecting information about research participants' beliefs, attitudes, interests or behavior through questionnaires, interviews, or pencil and paper tests" (Gall et al., 1999).

Teachers' Views. In this study, teachers' views refers to the perspectives, opinions, experiences and recommendations of teachers collected and analyzed by using survey method.

Validity. The appropriateness, meaningfulness and usefulness of the specific inferences made from a test score.

Limitations

Many policies have been enacted as part of Texas' educational reform effort; however, this study is limited to examining teacher perceptions and student outcomes during the last five years of high-stakes testing implementation. Additionally, the study focuses only on urban communities with high African-American and/or Hispanic populations.

The intent of the study is to place the school in a socio-historical context. There are several limitations associated with capturing the socio-historical context of a community. One is the limited documentation of African-American and Hispanic contributions to the community. The research frame of the study is limited to times spent within the community. Additionally, high-stakes assessment causes some teachers to feel pressure at their jobs. Texas superintendents, principals, and teachers find it hard to get jobs if they are from failing schools (Hurwitz & Hurwitz, 2000). Hence, opinions expressed by educators against a high-stakes testing and accountability system often are seen by the public as a "dodge" (Schmoker, 2002). This reluctance could result in non-responses or answers that truly do not reflect their views.

Delimitations

This research will not focus on school accountability, but only on assessments that have high stakes for individual students. The study is limited to perceptions of secondary school teachers and students in one central Texas school district, specifically focusing on high school students living within that district.

Assumptions

The study will be conducted in a district with a diverse student population, including African-American, Hispanic, and poor students. Teachers in the study are well-versed in high-stakes testing policies in the state of Texas. Additionally, these teachers understand the pedagogy required to teach diverse student populations.

Significance of the Study

The most basic obligation of educators at the secondary level is to meet the needs of students as they find them, with their different backgrounds, and to teach knowledge and skills to allow a student to grow and mature with meaningful expectations of a productive life in the workforce and elsewhere. During the last fifty years, assessment and accountability have played prominent roles in many of the reform efforts established to meet this goal. Testing and assessment have been the focus of controversy and the darlings of policymakers. Yet, after five decades of assessment and accountability, America's education system is not meeting the needs and expectations of many poor and minority students. According to McLaughlin and Shepard, the standards system has been corrupted, in many instances, into a heavy-handed system of

rewards and punishments without the capacity-building and professional development originally proposed as a part of the vision (McLaughlin & Shepard, 1995). Other researchers have pointed out that external accountability testing leads to the de-skilling and de-professionalization of teachers (Darling-Hammond, 1988). High-stakes testing teaches students that efforts in school should be in response to externally administered rewards and punishments rather than to the excitement of ideas. The 2001 Leave No Child Behind Program mandates that all students meet high expectations and standards, and states are using high-stakes testing as a means to implement educational reform.

High school exit exams are having a major impact on American students, especially minority students. Many of these students do not pass these exams and receive high school diplomas. In 2002, almost half of all public school students, and more than half of all minority students, lived in the eighteen states that require students to pass a test before graduation (Chudowsky et al., 2002). Six other states have begun to phase in exit exams, but have not yet started to withhold diplomas based on test scores. If states continue their current plans, by 2008, at least 24 states will have exit exams (Chudowsky et al., 2002).

The states with exit exams include several large or rapidly growing states, as well as states with higher-than-average minority populations. When the Center on Education Policy compared its exit exam data with the U. S. Department of Education enrollment data, they concluded that 49 percent of all public school students and more than 54 percent of all minority public school students currently live in states with exit exam requirements (*Digest of education statistics 2000*, 2001). By 2008, exit exams will affect approximately seven in ten public school

students and eight in ten minority students who currently live in states with an exit exam requirement.

A study of high school exit exams (Amrein & Berliner, 2002), observed that these tests are more common in states that:

Are located in the South and Southwest

- Have higher percentages of African-American and Hispanic students and lower percentages of white students than the national average
- Have greater degrees of poverty
- Allocate less school funding per pupil than the national average
- Have more centralized state governments, rather than powerful local governments
- Have high population growth compared to the nation.

Amrein and Berliner noted that many states with high school exit exams have lower levels of achievement, and there is always a correlation of low achievement to poorer students (Amrein & Berliner, 2002). Often, policymakers in these states pursued high-stakes testing more aggressively because of public concerns about low-performing schools.

The data do vary; not every state with a high percentage of minority students has adopted exit exams. However, Amrein and Berliner calculated that 75 percent of the states with an African-American population above the national average have exit exams. This percentage is projected to increase to 81 percent by 2008. Of states with Hispanic populations above the national average, 67 percent have exit exams; and by 2008, 89 percent will have them. Only 13

percent of the states with an above average of white students have implemented exit exams (Amrein & Berliner, 2002).

When state policymakers adopted high school exit exams, their goals were to monitor and increase student learning. They anticipated that students who make an effort would be able to meet the state standards, pass the tests, and receive diplomas. Policymakers were not prepared for the unintended consequences of high-stakes tests. Yet, the uncertainty principle warns, "The more important that any quantitative social indicator becomes in social decision-making, the more likely it will be to distort and corrupt the social process it is intended to monitor" (Amrein & Berliner, 2002, p.2). A study by Amrein and Berliner (2002) revealed that, if the intended goal of high-stakes testing policy is to increase student learning, then that policy is not working. While a state's high-stakes test may show increased scores, there is little support in these data that such increases are anything but the result of test preparation and/or the exclusion of students from the testing process. These distortions are predicted by the uncertainty principle.

This study will provide a better understanding of the significance and unintended impact of high-stakes assessments on children, especially poor, African-American, and Hispanic students. Darling-Hammond and Ancess (1996) have suggested that African-American and Hispanic populations are mostly adversely impacted by test-driven systems because they have higher failure rates and consequently a greater frequency of placements in lower-track classes (Soder, 1996). Other research confirms that, “the higher your family’s socio-economic status, the more likely you are to do well on a number of test items” (Popham, 1999). Students identified as minority, limited English-proficient, over-

age in grade, or economically disadvantaged, or labeled “at risk,” performed poorly on the TAAS (Montecel, Supik, & Cardenas, 1994). Although high school assessments were designed to promote student learning, unintended consequences of high-stakes testing have arisen with a reportedly profound negative effect on low-income, African-American and Hispanic students and their families. However, educators have limited knowledge of these unintended consequences. Thus, the present study has particular importance for communities with high numbers of low-income, African-American, and Hispanic students. It is hoped that the new data will allow educators a means to discuss the unintended consequences of high-stakes testing for which many students are unjustly being held accountable. Information found in this research hopefully will also enable other teachers and administrators to more readily equip themselves for the challenges and stresses found in assisting all students to successfully pass states’ exit exams and obtain high school diplomas. We must have dialogue and develop strategies that will allow us to make the Leave No Child Behind policy a reality for all children.

Summary

States often sell high-stakes testing as a tool to ensure accountability and a quick fix for education reform. Currently eighteen states, enrolling half of all public school students, require their students to pass exit examinations before they can graduate from high school and receive a regular diploma. High-stakes assessment is a crucial part of accountability. Many students begin high school which high expectations of successfully completing state high-stakes exams and graduating. However, many minority and poor students are not meeting this expectation.

Chapter I has presented an introduction to acknowledge high-stakes assessment as a key component of educational reform, has described factors that have stimulated the discussion on the unintended consequences of high-stakes assessment, and has addressed the need for and significance of this research study on unintended consequences of high-stakes testing. It has listed research questions to guide the study, has given a brief synopsis of methodology, and has described terms used in the research, as well as limitations, delimitation, and assumptions.

The purpose of this study is to determine the consequences of high-stakes testing on high school students, specifically those unintended side effects of high-stakes testing for economically disadvantaged and ethnic minority students from critical perspective of teachers and students. As clients of our educational system become more diverse in their characteristics, it becomes increasingly important to look at high-stakes testing and how it affects all children.

Chapter II reviews related literature. The first part of the chapter addresses the fact that this study will provide an historical and theoretical review of relevant studies on high-stakes testing. Chapter II also unpacks the assumptions and philosophical foundations upon which testing policies are based; it examines the pros and cons of high-stakes testing in the United States and Texas.

CHAPTER II – REVIEW OF THE LITERATURE

Accountability is the glue that holds together standards, curriculum, lessons, and assessments. It enables students to be successful learners. The purpose of this study is to determine the consequences of high-stakes testing on high school students, specifically, those unintended side effects of high-stakes testing for economically disadvantaged and ethnic minority students, from the critical perspective of teachers and students.

America has been involved in education for about 150 years. After 100 years, with the court case of *Brown v Topeka* in 1956, African-Americans and Hispanics obtained meaningful inclusion in the education process. During the next fifty years, minority groups were given various forms of universal access to education. Now, the Leave No Child Behind Act of 2002 requires that, within a 12-year period, by 2014, all students obtain universal proficiency.

President George W. Bush began this new era of universal proficiency in education. In the first week of his presidency, he proposed the same set of policies he had advocated while governor of Texas, including more reliance on standardized testing to judge schools. During the 2000 campaign, he highlighted Texas' reform record in education as evidence of his competence in education policy. However, the details of education policies focusing on testing are not definite in individual states. One issue in which debates over high-stakes testing has not settled down has been in the area of high school testing. In the past two years, the failure of a federal court case against the Texas graduation test requirements, the publication of one expert witness’s research on Texas, and two RAND Corporation studies that analyzed test scores from Texas, have increased

academic and civil-rights scrutiny of high-stakes testing as a gatekeeper for high school graduation ("GI Forum v Texas Education Agency," 2000; Grissmer, Flanagan, Kawata, & Williamson, 2000; Haney, 2000; Klein, 2000). Yet, increasingly, schools are administering tests that have important consequences for students.

The review of literature on high-stakes testing includes many theoretical writings and previous research dedicated to the examinations of the consequences of such policies on students and teachers. The review includes published and unpublished research, state-sponsored evaluation and research reports, papers from professional meetings, and articles from various organizations such as the Consortium for Policy Research in Education (CPRE).

In order to provide an appropriate review of high-stakes testing policy and its affect on teachers and students, this literature review begins with an historical presentation of education reform and testing policies. This is followed by an analysis of testing policies; theoretical arguments underpinning high-stakes testing policies; pros and cons of the high-stakes debate; review of previous research in the nation and Texas, including law, legal principles and actions; economics, student achievement and accountability; current high-stakes testing policies; high-stakes testing and curriculum; and students' outcomes, with a summary.

Historical Perspective: High-Stakes Testing in Secondary Schools

Assessment and accountability have played prominent roles in education. As statewide testing of students has become more and more common, researchers are examining the consequences of testing practices on students.

Today the practice of statewide assessment has been the center of a great deal of debate and political rhetoric (Popham, 1987; Shepard, 1992). Yet, the use of tests in the United States is not new. In 1845, Horace Mann promoted the administration of written essay exams (Madaus & O'Dwyer, 1999). He held a number of ideas requiring assessment that are remarkably similar to existing accountability policy. A historical review of high-stakes testing policies and procedures is required to fully understand the current high-stakes testing debate.

Assessment-Based Reform

Increasingly, schools are administering tests that have important consequences for students. At a micro-political level, families now make important decisions, such as where to live, based on test scores to rate neighborhood quality; and these scores affect property values. At the same time (macro-political level), in state and national levels, test scores are now commonly used to evaluate programs and allocate educational resources. Millions, perhaps billions, of dollars in resources now hinge on the tested performances of students in educational programs.

During five waves of educational reform tests, assessments have been the key. These waves include the role of tests in tracking and selection emphasized in the 1950s, the use of the tests for program accountability in the 1960s, minimum competency testing programs in the 1970s, school district accountability in the 1980s, and the standard-based accountability system of the 1990s (Linn, 2000).

Tracking and Selection – 1950s

High-stakes testing began with the launch of Sputnik in 1957. The Soviet Union, the United States' economic and political rival, beat the U.S. into space.

Interest groups, such as journalists and politicians, began to vigorously question American education (Kreitzer, Madaus, & Haney, 1989). At a macro-political level, state and federal politicians became more actively engaged in the conduct of education, including advocacy for the increased use of tests to assess schools and student learning. The writing of James B. Conant, in the 1950s (e.g., 1953), provided a rationale for "universal elementary education, comprehensive secondary, and highly selective meritocratic higher education" (Cremin, 1989). Tests were seen as important tools to support the implementation of Conant's conceptualization of the educational system, both for purposes of selecting students for higher education and for identifying students for gifted programs within comprehensive high schools.

Program Accountability – 1960s

In 1965, Congress adopted the Elementary and Secondary Education Act. The testing demands of the Title I Evaluation and Reporting System (TIERS) contributed to a substantial expansion in the use of norm-referenced tests (NRT) (Tallmadge & Wood, 1981). Rather than administering tests once a year in selected grades, TIERS encouraged the administration of tests in both the fall and the spring for Title I students in order to evaluate the progress of students participating in the program.

According to the National Association of Early Childhood Specialists in the State Department of Education (NAECS/SDE) (1991), Title I (previously called Chapter I) evaluation and practices helped "perpetuate inappropriate instructional methods and isolate Chapter I students from exciting and challenging experiences" (p.5), although Chapter I regulations permitted flexible criteria for the selection of children to be served and did not require the use of standardized

instruments for program evaluation until the beginning of the second grade. Most schools depended on large-scale, group-administered, norm-referenced tests for both selection of children and program evaluation. Many teachers and administrators observed that the over-reliance on group-administered, norm referenced tests "constrained their ability to move toward more developmentally appropriate practices in the kindergarten and primary years" (NAECS/SDE, 1995, P. 5), thus under-serving many students, including minorities.

Minimum-Competency Testing – 1970s

In the 1970s, a number of states adopted requirements under which students had to pass "minimum competency tests" as a condition of receiving high school diplomas, even if the student had satisfied all other requirements for graduation. As the name implies, the focus was on the lower end of the achievement distribution. Minimal basic skills, while not easy to define, were widely accepted as a reasonable requirement for high school graduation. The new requirements were of great importance for some students, but had little relevance for most students. Gains in student achievement were observed, but they occurred mostly at the low end of the distribution. Furthermore, some educators raised questions about the effects of generalizability on observed gains in test scores. Were students learning more, or were they becoming more astute about taking the test?

District Accountability – 1980s

The minimum competency test movement was almost entirely discarded. In 1983, the National Commission on Education report, "A Nation at Risk," called for an end to the minimum competency testing movement and the beginning of a high-stakes testing movement that would raise the

nation's standards of achievement. Accountability programs took a variety of forms, but shared the common characteristic that they increased real or perceived stakes of results for teachers and educational administrators.

The accountability systems of the 1980s relied heavily on published standardized tests. Upward trends in student achievement were reported by an overwhelming majority of states and districts during the first few years of accountability testing programs. However, a physician, John Cannell (1987), forcefully brought to public attention what came to be known as the Lake Wobegon effect (Koretz, 1988). This incredible finding was that essentially all states and most districts were reporting that their students were scoring above the national norm. Based on his review of the data, Cannell concluded that, "Standardized, nationally normed achievement tests give children, parents, school systems, legislatures, and the press inflated and misleading reports on achievement levels" (Cannell, 1987).

Researchers realized that, if student performance were improving nationwide, then comparison of results to old norms would put a current average performance above the mean of the old norms. However, gains on National Assessment of Educational Programs (NAEP) were more modest than the gains found on most standardized tests, which raised doubts about the generalizability or robustness of the putative gains that were reported on standardized tests (Linn, Grause, & Sanders, 1990).

The Lake Wobegon effect illustrated the repeated use of the same test form year after year, the exclusion of students from participation in accountability testing programs at a higher rate than they are excluded from norming studies,

and the narrow focusing of instruction on the skills and question types used on the test (e.g., (Koretz, 1988; Linn et al., 1990; Shepard, 1990).

High-Stakes Accountability – 1990s

In the late 1990s, the report, “A Nation Still At Risk,” warned of a rising tide of mediocrity in American public education; some states had replaced minimum competency tests with graduation exams measuring knowledge and skills at the tenth grade level or higher ("A nation still a risk: An education manifesto," 1998). Tests in accountability systems are not new. In the 1960s, payment for student performance results was the basis for holding school administrators, teachers and students accountable. Since that time, high-stakes accountability can be found cropping up and fading away over many decades. What is different about the current emphasis on performance-based accountability is its pervasiveness. "What is new is an increasing emphasis on student performance as the touchstone for state governance" (Elmore, Abelmann, & Fuhrman, 1996). Student achievement is not only being used to single out schools that require special assistance, but to provide cash incentives for improvements in performance. Furthermore, "The focus on performance has led to using outcome data, such as test scores and dropout rates, as criteria for accreditation" (Elmore et al., 1996).

The intent of this emphasis on outcomes is clear (Elmore et al., 1996), and is described in the underlying rationale as follows:

In principle, focusing on student performance should move states away from input regulations—judging schools based on the number of books in the library and the proportion of certified staff, for example—toward a model of steering by results—using rewards, sanctions, and assistance to move schools

toward higher levels of performance. In other words, the educational accountability should focus schools' attention less on compliance with rules and more on increasing learning for students (p. 65).

Whether focusing on accountability or student learning, currently, 23 states require students to pass graduation tests ("Making standards matter," 1999), increased from eighteen in 1998. Of the 23, fourteen now set graduation-test standards at the tenth-grade level or higher ("Making standards matter," 1999). Federal law takes no position on whether states and districts should use test results to determine whether individual students will receive high-school diplomas or be promoted to the next grade.

As a result of “A Nation Still at Risk,” state policymakers in every state but Iowa developed educational standards, and every state but Nebraska implemented assessment policies to check those standards (*Quality counts*, 2001). In fixing high stakes to assessments, policymakers and policy stakeholders borrowed principles from the business sector, and attached incentives to learning and sanctions to poor performance on tests. High-performing schools are rewarded.

No Child Left Behind Act – 2002

The No Child Left Behind Act (NCLB) reflects the growing consensus that all children can learn high standards if they are given the means to do so. The law rejects what President Bush has called “the soft bigotry of low expectations,” and requires progressively more extensive “corrective actions” in schools and districts in which students fail to satisfactorily progress (*Measuring what matters: An update on educational assessment and accountability*, 2002). NCLB law calls for statewide assessments in reading/language arts and mathematics (and

eventually science) in grades three through eight by the 2005-2006 school year (with science assessment), and specifies that states will develop an accountability system (Council of Chief School Administrators, 2002).

The accountability provisions of NCLB will affect education authorities at all three levels of government. Districts and states are held responsible for ensuring that districts make annual yearly progress (AYP). The federal government will oversee state progress in meeting the requirements of the new law (*Measuring what matters: An update on educational assessment and accountability*, 2002).

While the goals of the legislation are widely shared, No Child Left Behind raises unprecedented implementation challenges for states, districts and schools. In January 2002, *Education Week* determined that only nine states had assessment accountability systems that were likely in compliance with the law. The report noted pitfalls if these tools were misused (Olson, 2002). *Measuring What Matters*, a report by the Committee for Economic Development (2002), evolved to assist business leaders and other citizens in monitoring the implementation of NCLB. The report explores three challenges that must be met if "No Child Left Behind" is to accomplish its goal of improving student learning (*Measuring what matters: An update on educational assessment and accountability*, 2002):

- Challenge 1: Tests results must accurately measure and report student achievement.
- Challenge 2: Teaching must focus on important skills and knowledge, not test questions.

- Challenge 3: Low-performing schools must be given the help they need to improve.

Civil Rights activists and other stakeholders assert that America has not always believed that all children can learn. Furthermore, the criticism is made that administrators use tests that have negative consequences for students.

Effects of High-Stakes Testing on Students: A National Dilemma

Researchers are passionately debating whether high school exit exams are good or bad ideas. Research has mostly focused on three areas: student achievement, dropout rates and opportunities after high school. Many researchers and practitioners believe that standard-based reform and high-stakes testing will have their greatest effect on African-Americans, Hispanics, English-language learners, and low-SES students. However, there is serious dispute whether such testing will help or harm such students.

Arguments in Support of High-Stakes Testing

Proponents point out that African-American, Hispanic, and poor students are among those who are most poorly educated, and who therefore have the most to gain from a movement toward holding all schools accountable to high standards of teaching and learning. Over the years, this and other arguments have been used to promote high-stakes testing. A summary of these points included (Amrein & Berliner, 2002; Carnoy, 2001):

- students and teachers need high-stakes tests to know what is important to learn and to teach;

- teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to motivate the laziest ones to work harder;
- students work harder and learn more when they have to take high-stakes tests;
- students will be motivated to do their best and score well on high-stakes tests; and
- scoring well on the test will lead to feelings of success, while doing poorly on such tests will lead to increased efforts to learn.
- Supporters of high-stakes testing also assume that the tests:
- are good measures of the curricula that is taught to students in our schools;
- provide a kind of “level playing field,” an equal opportunity, for all students to demonstrate their knowledge;

Additionally, the supporters believe that:

- teachers use the test results to provide better instruction for their students;
- administrators use the tests to design better professional development for teachers; and
- parents understand high-stakes testing and how to improve student scores.

Critics of high-stakes testing

The validity of these statements in support of high-stakes testing has been examined through both quantitative and qualitative research, and by the commentary of teachers who work in high-stakes testing environments. And in

fact, some research studies show exactly the opposite of the effects anticipated by supporters of high-stakes testing (McNeil & Valenzuela, 2000; Orfield & Krohaber, 2001; Sacks, 1999; Sheldon & Biddle, 1998).

Another possible consequence of high school high-stakes testing is an increased number of high school dropouts. Gary Natriello agrees with this view. In fact, he argues that testing is perceived to be an efficient way of assessing school quality, but also that it can be used as a barrier to graduation or educational advancement ("The impact of high-stakes testing policies or minority and disadvantaged students," 2000). According to Gary Orfield, high-stakes testing hurts low-income and ethnic minority students, and is linked to high dropout rates among these groups; also, African-Americans and Hispanics are three to four times more likely to be retained than whites. “States should know who is being hurt by these tests,” says Orfield. He warns that tests are not standards, but they can be the punishment of innocent victims of unequal education” ("The impact of high-stakes testing policies or minority and disadvantaged students," 2000). Many believe that higher standards—or the expectation of eventual failure—will accelerate decisions to leave school on the part of marginal students. Some argue that high standards create pressure on school administrators to have certain students leave school early if they have poor chances for graduation (Haney, 2000).

Critics of high-stakes tests assert that high-stakes testing does not promote real learning. Also, they fear many students will be harmed by high-stakes testing: students will be retained in grades or denied diplomas because their school did not expose them to the skills and learning need to pass the test (Heubert & Hauser, 2001). They argue that high-stakes testing does not promote

real improvement in learning. Rather, teachers and principals are motivated to meet the standards by teaching to the test. Meyer has argued that in a high-stakes accountability system, teachers and administrators are likely to exploit all avenues to improve measured performance. For example, teachers may "teach narrowly to the test. "For tests that are relatively immune to this type of corruption, teaching to the test could induce teachers and administrators to adopt new curriculums and teaching techniques much more rapidly than they otherwise would (Meyer, 1996, p. 140).

Instead of creating an improved learning environment, these crude forms of assessment may reduce opportunities to higher learning skills, particularly for low-income students (McNeil & Valenzuela, 2000). Linda McNeil and Angela Valenzuela further maintain:

The pressure to raise TAAS scores leads teachers to substitute commercial TAAS-prep materials for the substance of the curriculum . . . Subjects tested by TAAS (reading, writing, and mathematics) are reduced, in the test and test-prep materials, to isolated facts and fragments of facts. This artificial treatment of these isolated components may enable children to recognize the components on a multiple choice test, but not necessarily enable them to use the components in other contexts (McNeil & Valenzuela, 2000).

Critics also assert that state testing increases the probability of disadvantaged students dropping out of school by forcing students to repeat grades (Haney, 1999, 2000; Shrag, 2000).

Even on tests that measure basic skills, minority students fail at a higher rate than other students, especially in years after such tests are first introduced. For instance, in the 1970s, when minimum-competency tests gained popularity,

twenty percent of the African-American students compared to two percent of white students failed Florida's graduation test and were denied graduation diplomas (*Debra P. v Turlington*, 1979). This also appears to be true for a graduation test recently adopted. In Texas, for example, which based its graduation test on eighth-grade levels, Shrag reports that the pass rate of African-Americans and Hispanics doubled between 1994 and 1998 (Shrag, 2000); and that failure rates of African-Americans and Hispanics narrowed during that time (Viadero, 2000). Regardless, 1998 data from Texas graduation tests shows continuing disparities: cumulative failure rates of 17.6 percent for African-Americans, 17.4 percent for Hispanics, and 6.7 percent for Anglo students (Natrellio & Pallas, 1999). Although these students are not officially listed as dropouts, they are. They have been pushed or tested out of the system without a high school diploma after completing all courses required by their schools.

Texas and other states are moving to higher-standard-based graduation exams. Based on National Assessment for Educational Progress (NAEP) data, about 38 percent of all students would fail tests that reflect world-class standards if they were administered today (Linn, 2000); an 80 percent failure rate is projected for minority and English-language learners on tests that embody world-class standards. These predictions are consistent with recent data from Massachusetts, where students have begun taking graduation tests that reflect world-class standards (Heubert, 2000). Large-scale promotion testing, which is especially pronounced in large urban school districts ("Making standards matter," 1999), has led to very high grade retention for African-American, Hispanic, and English-language learners. The strongest predictor of whether students will drop out of school is whether they have been retained in grade. With the rapid growth

of promotion testing, a disproportionate number of African-Americans and Hispanics will be at an increased risk of dropping out of school by virtue of having been retained in grade one or more times. Students who were retained in grade are much more likely to drop out of school than students who were not retained (Hauser, 1999; Shepard & Smith, 1989). It is also likely to reduce the number of students who remain in school long enough to take the graduation test. Many also believe that some minority students are being forced out of schools and that their legal rights are being violated, as discussed in the next section.

Laws, Legal Principles, and Actions

Beliefs and Perceptions

Tests, their analysis, and perceptions of ethnic groups and/or special populations, are embedded in the educational fabric of the United States. To understand why laws and court cases have been motivated in America over the testing of the nation's children, it is important to look at the historical perspective and attitudes the country has expressed about the abilities of different ethnic groups. The basic underlying assumption of tests in schools in the early 1900s viewed talent, ability, and intelligence as products of heredity, and was often used to explain the differences between ethnic groups. Social Darwinism, the Binet test, and the Army Alpha Test helped form these concepts and beliefs held by some Americans about testing and ethnic groups.

Social Darwinism was the educational system's response to the diverse student population. Social Darwinism, a scientific theory that was viewed as a scientific law for viewing groups of people as less socially and economically developed than others, held that ethnic minorities were evolutionarily inferior to

white Protestants. Social Darwinism provided states, communities, and schools with the “scientific” justification to treat children of various ethnic groups differently, and tests administered in this period validated the practice of grouping by race and ethnicity.

In 1910, the Binet intelligence test was introduced to American schools. The Binet test was a popular testing tool used to screen immigrants; often, immigrant children were classified as feeble-minded. In the 1920s, schools discovered pencil and paper tests. A board intelligence-testing program developed by the US Army influenced these assessment procedures. A new group of pencil and paper tests, consisting of multiple-choice and true/false questions that could be scored with a key, were developed. Studies indicated the testing results were correlated to supervisors’ judgments and ethnic backgrounds (Widgor & Garner, 1982), leading to a massive study conducted by the National Academy of Sciences (Yerkes, 1921), which analyzed 160,000 recruits. This study found that non-immigrant whites scored the highest on the tests, while the lowest-scoring recruits were from the Southern and Eastern European countries.

By the 1950s and early 1960s, the belief that ability to learn was determined genetically and hereditarily was challenged by an environmentalist view. General belief had shifted such that talent and intellectual ability were seen as being distributed across racial and ethnic groups. According to (Widgor & Garner, 1982), “Within this frame of reference, which came to be pretty much the conventional wisdom by the 1950s, testing seemed a liberating tool that could circumvent the privileges of birth and open the doors to opportunity to Americans of all kinds. Thus opening doors for laws and policies against discrimination.”

Nondiscrimination Laws

High-stakes testing and accountability are embedded in four nondiscrimination laws that have been enacted by Congress: Title IV of the Civil Rights Act of 1964 (Title IV), Title IX of the Education Amendments of 1972 (Title IX), Section 504 of the Rehabilitation Act of 1973 (Section 504), and Title II of the Americans with Disabilities Act of 1990 (Title II). Title IV, Title IX, Section 504, and Title II, as well as the equal protection clause of the Fourteenth Amendment of the U.S. Constitution, prohibit intentional discrimination policies based on race, national origin, sex, or disability, or practices that have a discriminatory and disparate impact on students based on their race, national origin (including limited English proficiency), sex, or disability. The Section 504 regulation and the Individuals with Disabilities Education Act (IDEA) contain specific provisions relevant to the use of high-stakes tests for individuals with disabilities (U. S. Department of Education, Office of Civil Rights, 2000). The Office of Civil Rights highlights some of the issues that have been considered by federal courts in assessing the legality of specific testing practices for making high-stakes decisions which include:

- The use of educational tests for the purpose of which the test was not designed or validated;
- The use of test scores as the sole criterion for the educational decision;
- The effects and nature of the opportunity provided to students to master required content, including whether classroom instruction included the material covered by a test administered to determine student achievement;

- The significance of any fairness problems identified, including evidence of differential prediction criteria and possible cultural biases in the test or in test items; and
- The educational basis for establishing passing or cutoff scores.

The federal nondiscrimination regulations also provide that a recipient of federal funds may not "utilize criteria or methods of administration which have the effect of subjecting individuals to discrimination" (U. S. Department of Education, Office of Civil Rights, 2000 , p. 15).

Courts applying the disparate impact tests have generally proceeded to examine three questions to determine if the practice at issue is discriminatory:

- Does the practice or procedure in question result in significant differences in the award of benefits or services based on race, national origin, or sex?
- Is the practice or procedure educationally justified (educationally necessary)?
- Is there an equally effective alternative that can accomplish the institution's educational goals with less disparity?

And in due process cases, federal courts have required, as a matter of "fundamental fairness," that students have a reasonable opportunity to learn the materials covered by the test where passing the test is a condition of receipt of a high school diploma or a condition of grade-to-grade promotion (*Brookhart v Illinois Board of Education*, 1983; *Debra P. v Turlington*, 474 F Supp 244 (M D Fla. 1979), 1979; *GI Forum v Texas Education Agency*, 2000).

Constitutional Issues

Constitutional challenges to testing programs under the Fourteenth Amendment have raised both equal protection and due process claims. The equal protection principles involved in discrimination cases are the same as the standards applied to intentional discrimination claims under the applicable federal nondiscrimination statutes. The due process clause of the Fourteenth Amendment is particularly associated with cases challenging the adequacy of the notice provided to students prior to this type of test, and the students' opportunity to learn the required content.

Court Rulings for High-Stakes Testing

Allegations of racial bias continue to plague high-stakes tests today. Court rulings on the constitutionality of high school exit exams (e.g., *Debra P v. Turlington*, 1981, 1984; and *G. I. Forum et al. v. Texas Education Agency*, 2000) have identified specific criteria for educational agencies to use when determining if a high school student has had the opportunity to learn. These criteria include multiple opportunities to take the test; students having been taught the tested skill; adequate notice of graduation testing requirements; and opportunities for successful remediation.

There are several court cases that affect the implementation of high-stakes tests, including those highlighted below:

- **1997 – *G. I. Forum and Image de Tejas v Texas Education Agency***. The plaintiffs argued that Texas's state test had a disparate impact against African-Americans and Hispanic students. The judge did not deny that a large gap existed in the pass rate between Anglo and minority students, but ruled in favor of the

defendants. The judge concluded that the state had demonstrated "education necessity"; simply put, the high-stakes test approach was the only way to force improved outcomes in Texas education.

- **2001** ***Rene v Reed***. The Plaintiff argued that she had not been exposed to the curriculum tested on Indiana's high school exit exam. The Court of Appeals of Indiana found that the state requirements to provide remedial assistance to all students who did not meet academic standards made it implausible that the students were not exposed to the subject tests throughout their high school career.

This analysis focuses on two criteria which are often the most difficult for states to achieve: (1) testing what students have been taught; and (2) support for remediation. Exit exams should and must test what students have been taught. While most people would agree with this principle, how do we ensure that more than 22,000 public schools and 1.2 million public school teachers are testing, and 13.5 million public high school students are being tested in, what is being taught (Nation Center for Education Statistics, 2002)? To measure the opportunity to learn, states must look at their formal standards and curriculum to see if they include knowledge and skills that the test is designed to measure; that resources and materials cover the content; but, most importantly, that the actual curriculum and instruction is being delivered in the classroom. It is hard to judge curriculum delivery and instruction in the classroom. Many schools with high African-American and Hispanic populations often are staffed with new or inexperienced teachers, out-of-field teachers, or teachers who have proven to be ineffective at

other schools. Most states have not taken steps to ensure that equitable instruction occurs across the board (Center on Education Policy, 2002).

An important part of the opportunity to learn is providing students who fail exit exams with instructional support to help them succeed the next time they take the test. Many states have policies that require local districts to provide some type of remediation to students who fail the test on the first attempt. According to the Center on Education Policy (2002), professional development is a critical part of mounting an effective remediation program. Some states provide funding and leadership, professional development, and high quality materials to local districts. Only half of the states with mandatory graduation tests also allocate state funds for instructional assistance to students who fail exit exams ("Hot topics: Assessment," 2002). However, what do economic studies reveal about minorities, education, and funds? White America frequently asks, “Is it worth the cost to educate historically disenfranchised children?”

Economics, Student Achievement, and Accountability

Today’s question in education is how best to improve student achievement. Today’s public school reformers are calling for greater productivity in our schools.

Summary of Major Ideas of Resources and Student Achievement

Until recently, many social scientists thought that providing schools with additional resources would have little impact on student achievement, the so-called “money doesn’t matter” thesis (Ladd, 1996). This view dates from the “Coleman Report,” one of the earliest product-function studies, which found family influence to be strong and school resources of little effect of (Coleman,

1966). Eric Hanushek (Hanushek, 1989, 1994, 1996) also argues that evidence from over 300 empirical studies provided no consistent evidence that increased school resources raised achievement scores.

In the early 1990s, Hedges and colleagues disagreed with Hanushek, and conducted a formal meta-analysis that Hanushek had reviewed. Hanushek found that most of these studies lacked statistical power to detect resource effects even when the samples were quite large. When Hedges and his colleagues polled data from all available studies, the result indicated positive, statistically significant evidence that some programs may have large effects (Hedges & Greenwald, 1996). Additionally, other works conducted with alternate methodologies, like Hierarchical Linear Modeling rather than the production function framework, often showed positive effects of resources (Grissmer & Flanagan, 1998).

Greenwald, Hedges, and Lane (1996) used meta-analytic methods to assess studies that reported the direction and magnitude of the relations between a variety of school inputs and student achievement. The research studies aggregated data at the level of school districts or smaller units, and either controlled for socioeconomic characteristics or were longitudinal in design. The analysis found that a broad range of resources was positively related to student outcomes, with effect sizes large enough to suggest that moderate increases in spending may be associated with significant increases in achievement.

Accumulating evidence is now challenging both NAEP evidence and the accuracy of previous empirical studies. According to Grissmer and his colleagues, there is sufficient evidence to replace the “money doesn’t matter” hypothesis with one that asserts that money does matter for students from less

advantaged backgrounds, and minority students, but may not matter for students from more highly advantaged backgrounds (Grissmer et al., 2000).

Current High-Stakes Testing Practices

According to the Center on Education Policy (2002), twenty-two states offer incentives for high or improved test scores. Twenty states distribute financial rewards to successful schools, and nineteen states distribute rewards to improved schools. Punishments are attached to school scores twice as often as rewards. Forty-five states hold schools accountable for test scores by publishing school and district report cards. Twenty-seven of those states hold schools accountable through rating and ranking mechanisms; fourteen have the power to close, reconstitute, or take over low-performing schools; sixteen have the authority to replace teachers and/or administrators; and eleven have the authority to revoke a school's accreditation.

For administrators, the threats of termination and cuts in pay exist, as does the potential for personal bonuses. For example, in Oakland, California, city school administrators can receive a nine percent increase in pay for good school performance with the potential for an additional three percent increase.

Teachers with low average class scores may be prevented from receiving salary increases; also, performance may influence tenure decisions, and, in sixteen states, may cause dismissal. Texas has even linked teacher evaluations to student or school test results, and more states plan to do this in the future. Schools and teachers are not the only targets of rewards and punishments for test performance. Thus, educators and other stakeholders are concerned about the characteristics of such exams.

Salient characteristics of high-stakes testing indicate that use of large-scale achievement tests as instruments of educational policy is becoming a national norm. States and schools are making high-stakes decisions with important consequences for individual students. Three such high-stakes decisions involve tracking, promotion to the next grade, and determining whether a student will receive a diploma at graduation. Analyses of data show that high school graduation exams are more common in states that allocated less money than the national average per pupil for schooling, as compared to the nation (*Digest of education statistics 2000*, 2001), and are more likely to be found in states that have a more centralized government, rather than those run by strong county or city governments (Elazars, 1984); they are also most likely to be found in the southwest and the south. High school graduation exams are currently in use in fifty percent of the southwestern states and 66 percent of the southern states (Amrein & Berliner, 2002). A high-stakes testing policy in connection with the graduation requirement is more likely found in states with higher percentages of African-Americans and Hispanics and a lower percentage of whites as compared to the nation as a whole.

Seventy-five percent of the states with a higher percentage of African-Americans than the national average have high school graduation exams. According to Amrein and Berlin, by 2008, 81 percent of such states will have implemented high school graduation exams. Sixty-seven percent of the states with a higher percentage of Hispanics than the national average will have high school graduation exams. Simply put, high-stakes exams affect students from racial minority backgrounds in greater proportions than they do white students. High school graduation exams disproportionably affect students from lower

socioeconomic backgrounds. High school graduation exams are likely to be found in states with the greatest degree of poverty (Amrein & Berliner, 2002). Matters of national standards and implementation of high-stakes tests are less likely to be concerned with the reform of relatively elite schools (Ohanian, 1999).

High-Stakes Testing and Curriculum

Most research on the effects of high-stakes testing has looked at curriculum and instruction. Standards-based reform has sought to quiet critics of minimum competency testing by raising expectations for student learning beyond basic skills to more demanding levels (Chudowsky et al., 2002). Yet, many states do not have alignment between assessments, standards, and instruction, thus disproportionately affecting students, especially minority students' abilities to meet the standards and successfully pass exams. More states are moving toward standards. Standards provide a more defensible basis for high-stakes exams by defining the content that students should learn by the end of high school. In the 1980s, states had not yet developed academic content standards, so standards were merely implied by the content of the prior forms of minimum competency exam.

Several claims have been made as to the possible positive and negative effects of high-stakes testing on curriculum. Popham believes that curriculum will improve as schools, teachers and students attempt to meet the challenges testing will impose (Popham, 1987). Madaus & O'Dwyer (1988) and Shepard (1992) fear high-stakes testing will narrow curriculum, focus on low-order skills, and/or take control of the curriculum away from local sources.

In a study comparing Pennsylvania, a low-stakes testing state, with Maryland, a high-stakes testing state, teachers in the high-stakes testing solution

reported a narrowing of the curriculum (Corbett & Wilson, 1990). But all teachers did not feel this was a bad thing. They stated, “Maryland school districts focused more directly on improving test scores, altered the curriculum to a greater extent, reported more improvement in the curriculum, and felt the curriculum had narrowed more than their Pennsylvania colleagues” (Corbett & Wilson, 1990,p. 72). The teachers did not always think curriculum changes were in the best interest of the students.

A five-year study, involving 12,404 eleventh grade students from the Austin Independent School District, conducted by Rodgers, Paredes and Mangino (1991), looked at the effects of the Texas Educational Assessment of Minimum Skills (TEAMS), a test that students needed to pass in order to graduate from high school. The test focused on language arts and math (Rodgers et al., 1991). Rodgers and colleagues found that basic skills, as measured on the Tests of Achievement and Proficiency (TAP), increased as a result of the minimum-competency exam. However, higher-order thinking skills remained the same. The researchers concluded that districts should be cautious about narrowing the curriculum and letting higher-order thinking skills suffer for the sake of improving test scores.

Using qualitative interviews in a high-stakes elementary setting, Rottenberg and Smith used the Iowa Test of Basic Skills (ITBS) to evaluate curriculum decisions made by principals (Smith & Rottenberg, 1991). They found that testing reduced the time available for ordinary instruction. Schools were also neglecting materials not in the tests, while encouraging the use of instructional methods resembling testing—for example, multiple-choice exams.

In the study, teachers spent four weeks on intensive test preparation, plus two weeks administering the test itself. This emphasis on preparation was not limited to the time surrounding the test administration: 68 percent of the teachers reported using worksheets through the year to review expected test questions.

The literature provides very little evidence about the effects of high-stakes testing on the curriculum. However, the results are consistent in showing that high-stakes testing does affect how teachers teach. And these studies raise questions about high-stakes test implementation, student proficiency, and outcomes.

Student Proficiency and Outcomes

Student Learning and High-Stakes Testing

Researchers do not agree on whether or not high school high-stakes exams increase student learning. When African-Americans and Hispanics protest that high-stakes testing does not improve student learning, their remarks are often seen in a negative light. They may be accused of playing the “race card,” and their concerns are frequently disregarded as an excuse from people who cannot compete with whites.

Often, when a new mandatory high-stakes tests is introduced, such as a graduation test, student performance on that test increases for the first few years (Linn, 2000). The key question is whether other indicators of student learning corroborate gains on high-stakes exams. A majority of studies have found no evidence that exit exams increase student learning, as measured by other indicators such as standardized tests (Jacob, 2001; Neill & Gayler, 1999).

Amrein and Berliner Analysis

Amrein and Berliner (2002) conducted a “time series analysis” for eighteen states that used graduation exams for high-stakes purposes. They looked at the changes in student performance on four different standardized tests—SAT, ACT, Advanced Placement (AP) exams, and the NAEP—focusing on whether average scores on these other measures rose or fell in the years after the state began using the exam for high-stakes purposes. Although some states showed increased scores on their high-stakes exams over time, the researchers reported that these gains typically did not transfer to the other measures.

Amrein and Berliner examined K-12 high-stakes testing policies in eighteen states (Alabama, Florida, Indiana, Louisiana, Maryland, Minnesota, Mississippi, Nevada, New Jersey, New Mexico, New York, North Carolina, Ohio, South Carolina, Tennessee, Texas, and Virginia), inquiring whether testing programs promoted the transfer of learning that they were intended to foster. These states have the most severe consequences written into their K-12 testing policies, and lead the nation in incidences of school closures, school interventions, state takeovers, and teacher/administrator dismissal. They have the most severe K-8 promotion and retention policies and high school graduation exam policies. Additionally, they are the only states in which students are being denied regular high school diplomas, or are simply not graduating, because they have not passed the state high school graduation exam.

In the Amrein and Berliner study, the effects of high-stakes tests on learning were measured by examining indicators of the students learning, academic accomplishment, and achievement other than the test associated with high-stakes. The four different measures used were ACT, SAT, NAEP and AP

exams. If high-stakes testing causes teachers to upgrade curricula and instruction, and motivates students to study harder, then scores should also increase on other independent assessments (Neill & Gayler, 1999).

ACT Data

Sixty–seven percent of the states that use high school graduation exams posted decreases in ACT performance after high school graduation exams were implemented. On average, as measured by ACT, college-bound students decreased in levels of academic achievement (Amrein & Berliner, 2002). If the participation rate in the ACT program serves as an indicator of motivation to attend college, then one can conclude that most students taking the test were motivated to do their best

SAT Data

Fifty-six percent of the states that use high-stakes high school graduation exams posted decreases in SAT performance after those exams were implemented. Eight states displayed positive effects, while ten states displayed overall negative effects. But the gains or losses in scores were related to increases and decreases in the percentage of students participating in the SAT. Thus, it is hard to attribute any effects on the SAT to the implementation of high-stakes testing. Gains and losses in SAT scores are more strongly correlated to who participates in the SAT than to the implementation of the high school graduation exams. SAT participation rates fell 61 percent in those states with high-stakes exit exams (Amrein & Berliner, 2002). These participation rates are unsupportive of the belief that high-stakes exams will prepare and motivate more students to attend college.

NAEP Data

Some researchers argue that using ACT and SAT scores to assess the effects of high school graduation exams is illogical because high school graduation exams are designed to raise the achievement level of those students who are most likely to fail—the poor and racial minorities. These students do not take ACT or SAT in large numbers. However, most researchers agree that the effect of high-stakes policies can be assessed using the "national report card," NAEP. In general, there has been some improvement in the NAEP scores. This is particularly interesting because these slight improvements have occurred during a period of demographic change. African-American and Hispanic children (the subgroups with the lowest average scores) constitute a greater percentage of the school-age population than the NAEP-taking population thirty years ago (Amrein & Berliner, 2002). When greater shares of the test-takers come from lower-scoring subgroups, the national average is depressed. In other words, the overall gains are less than they would have been if the demographics composition of the test-takers had remained the same. In some subjects, as explained by Pascal D. Forgione, the overall average gain, paradoxically, may be lower than the gains for any one subgroup (Forgione, 1999).

When looking at NAEP data, it is important to understand who is actually taking the test. NAEP uses a multi-stage random sampling technique; each state and participating school district is randomly sampled. Once the school has been selected, school personnel look at their lists of students and remove any student classified as Limited English Proficient (LEP) and those students who have Individualized Education Plans (IEPs) as a part of their special education programs. It is well documented that a large number of Hispanics for whom

English is a second language and a disproportionate number of African-American students are placed in special education programs, eliminating large numbers of Hispanic and African-American students (Amrein & Berliner, 2002). Thus, gains and losses in NAEP data are more related to who was excluded from the NAEP than to the effects of high-stakes testing programs in a state.

Haney found that the exclusion rates explained gains in the NAEP scores. Texas, for example, was one of the states in which large gains in the NAEP scores were heralded as proof that high-stakes tests do indeed improve student achievement (Haney, 2000). However, the exclusion rate increased in grade four from eight percent to eleven percent, and in grade eight from seven percent to eight percent, from 1992–1996, prompting Haney to term the scores gained in Texas as "illusion arising from exclusion."

NAEP mathematics data indicated that high-stakes testing policies did not generally improve the performance of students on the grade four math test. High-stakes testing policies did not consistently improve the performance of students on the grade eight NAEP math test.

High-stakes testing policies did not consistently improve the general learning and competencies of students who took the fourth grade NAEP reading test from 1992–1998. By 1998, 75 percent of the states in the study had exclusion rates higher than the national average (Amrein & Berliner, 2002).

AP Data

1995–2000, the percentages of students passing various AP exams do not indicate that high-stakes high school graduation exams improve student achievement (Amrein & Berliner, 2002). Students' performance on the ACT, SAT, NAEP and AP exams raise questions about the validity of high-stakes tests as an

indicator to improve student learning and achievement, as well as the possible failure of the high-stakes policy initiative. If failure in attaining the goals for which the policy was created results in disproportionate negative effects on the life chances of America's poor and minorities, as it appears to do, then high-stakes testing policy is more than a benign error in political judgment (Amrein & Berliner, 2002), or merely with unintended consequences. It is an error in policy that is causing structural and institutional activities that discriminate against African-American, Hispanic, and poor students. Hence, there is an immediate need to investigate the unintended consequences of high-stakes testing. One of these consequences is a suspected negative relationship between high-stakes testing student dropout rates.

Impact Of High-Stakes Testing and Student Dropout Rates

The relationship between exit exams and dropout rates has been studied extensively. In 1989, a study found that states with the highest dropout rates had minimum-competency exit exams, while none of the states with the lowest dropout rates used state tests for graduation or other high-stakes purposes (Kreitzer et al., 1989). This does not necessarily mean that the high-stakes exams cause more students to drop out. Some limited evidence of a causal connection between exit exams and rising dropout rates comes from James Cattarall's 1989 study that includes the interview of 700 students from four states with exit exams. Cattarall found that students who initially failed high-stakes exit exams are more likely to express doubt about finishing school (Cattarall, 1989). Some studies support charges that exit exams have a more negative impact on the lowest-achieving students. Jacob found that low-achieving students in states

with exit exams were about 25 percent more likely to drop out of school than comparable peers in states without exams (Jacob, 2001).

Research also has linked failure on exit exams with increased dropout rates among high-achieving minority students. Griffin and Heidorn collected data from 76,000 Florida students, and concluded that students who fail the state graduation test were more likely to drop out of school than other students—but only if the student had a relatively high grade point average (Griffin & Heidorn, 1996). Students with lower GPAs did not seem to be affected by the test. The researchers speculated that the stigma attached to failing the test might cause students with good academic records to experience an acute drop in self-esteem or sense of embarrassment (Griffin & Heidorn, 1996).

Texas has had a statewide high school graduation test since 1986, Texas Educational Assessment of Minimum Skills (TEAMS) (until recently, Texas Assessment of Academic Skills, or TAAS). Walt Haney looked at a school's completion rates (basically the opposite of graduation rates over a twenty-year period) (Haney, 2001). He calculated the rates by dividing the number of high school graduates in a given year by the number of ninth-graders three years earlier. His data showed that, for Anglo students, the ratio of on-time graduates fell slightly between school years 1978-1979 and 1990-1991. But in 1991-1992, the first year after implementation of the new TAAS exam, the ratio showed the steepest drop in the entire twenty-year period. His data on African-American and Hispanic students is most upsetting and troubling. There is a severe decline in the number of minority students graduating from high school. Like their Anglo counterparts, the ratio of on-time graduates for minority students fell slightly in 1978-1979. They also experienced a steep decline in graduation rates in 1990-

1991. The graduation rates of minorities recovered slightly in 1992-1993. However, minority students have never reached the higher graduation level of the pre-1991 era. Texas has implemented a new exam, Texas Assessment and Skills (TAKS). The Texas State Board of Education has predicted a 65 percent failure rate for minority students. Those students who fail the exit exam will not receive a high school diploma.

Several other factors in conjunction with exit exams have been shown to be associated with dropping out of school (*Understanding dropouts: Statistics, strategies and high-stakes testing*, 2001). Students who have been retained in grade, who have poor grades and low-test scores, are most likely to drop out later on. Hispanic students are most likely to drop out, followed by African-American students and then Anglo students. According to the Civil Rights Project (*Dropouts keeping students in schools*, 2002), despite the importance of high school education, completion rates over the past 25 years have only increased slightly. In 1993, approximately 381,000 students in grades ten through twelve dropped out of school; and approximately 3.4 million persons in the United States, ages 16 through 24, were high school dropouts. Current research suggests that dropping out of school is a process rather than an event. A negative school experience influences many students to leave before they receive their diplomas, and one major cause is high-stakes testing and grade retention (*Dropouts keeping students in schools*, 2002).

Additional research on the impact of high-stakes testing and dropout rates is indicated. In at least seven states that provide dropout data to the federal government, minorities make up over 50 percent of the number of dropouts (*Dropouts keeping students in schools*, 2002). Currently, research shows that

dropouts are more likely to be unemployed than those who complete high school. When dropouts do secure work, on average, they earn less money than their peers who have high school diplomas. Dropouts are more likely to receive public assistance. And 68 percent of all prison inmates are high school dropouts (*Dropouts keeping students in schools*, 2002). Policymakers and researchers must continue monitoring the consequences of high-stakes testing.

Implementation of High-Stakes Testing

From a culturalist standpoint, the knowledge that every child benefits from high expectations and standards is not the problem for African-American and Hispanic parents. However, they are concerned with how their children are being taught and the lack of culturally relevant instruction in their classrooms. Culturalists are aware of the negative effects brought about by not seeing one history, culture or background in textbooks or curriculum, or by seeing that history, culture, or background distorted. Hence, they wonder how a non-minority teacher looks out into a classroom and sees the sons and daughters of slaves and migrant workers. How does this vision translate into his/her vision for educational excellence (Ladson-Billings, 1994) via class practices and assessment? They are especially concerned about teachers who say that they are color–blind, for claiming not to notice a child's color or ethnicity is dismissing the most salient features of a child's identity; and the teacher does not account for it in his/her curriculum planning, instruction, and assessment (Ladson-Billings, 1994). Hence, the teacher is limited in his/her ability to ensure all students have an equal opportunity to learn.

Thus, the implementation of high-stakes testing is an issue. And ensuring that every child has the opportunity to learn the material being tested is at the

heart of the debate over exit exams. This involves providing students with the full range of instructional services and support that is essential to meet the standards: well-qualified teachers, effective instruction aligned with state standards; high quality instructional materials and equipment; and extra learning time in a culturally relevant context for the student.

Validity

In using tests as a part of the high-stakes decision-making process for graduation or grade promotion, educational institutions, according to the U.S. Department of Education, Office of Civil Rights, should ensure that the test will provide accurate results that are valid, reliable, and fair for all test-takers. Test validity refers to a determination of how well a test actually measures what it says it measures. There are three central points to keep in mind regarding validity:

- The focus of validity is not really on the test itself, but on the validity of the inferences drawn from the results for a given use;
- All validity is really a form of “construct validity”; and
- In validating the inference of the test result, it is important to consider the consequences of the test’s interpretation and use.

When schools, districts, states, or other authorities mandate educational testing programs, the ways in which test results are intended to be used should be clearly described. It is the responsibility of those who mandate the use of tests to monitor their impact, and to identify and minimize potential negative consequences. Consequences resulting from the use of the test, both intended and unintended, should also be examined by the test user (Council of Chief School Administrators, 2002).

Reliability

Reliability refers to the degree of consistency of test results, test administrations, forms, items, scorers, and/or other facets of testing. No test or testing instrument is ever "error free." Thus, all indices of reliability are estimates of consistency, and all the estimates contain some error. Consistency over parallel forms of a test occurs when forms are developed to be equivalent in content and technical characteristics (Council of Chief State School Administrators, 2002).

Fairness

Tests are fair when they yield score interpretations that are valid and reliable for all groups of students who take the tests. The tests must measure the same academic construct (knowledge and skills) for all groups of students who take them, regardless of race, national origin, gender or disability. Culturally biased tests should not be used for the placement of African-American and Hispanic youth because they do not reflect the true ability of many students. Many tests, such as the Stanford-Binet Intelligence Test, include items that assess moral opinions and other values that reflect social bias rather than ability (Hillard, 1976). Consequences of inappropriate tests include (Kuykendall, 1992):

- the over-representation of African-American and Hispanics in special education classes and low-ability groups and tracks;
- the alienation and physical or psychological withdrawal of underachieving students from the learning process due to their inability to master mainstream culture well enough to do well on culturally biased exams;

- negative attitudes toward the schools on the part of parents who have come to recognize and appreciate the social skills of their children and who themselves felt victimized by culturally biased tests when they were in school; and
- lower expectations of teachers who fail to understand that lower scores of some African-American and Hispanic youths on tests is more an indication of cultural conflict than low intelligence.

According to Heubert and Hauser, fairness, like validity, cannot be properly addressed as an afterthought. It must be confronted throughout the interconnected phases of the testing process, from test design and development to administration, scoring, interpretation, and use (Heubert & Hauser, 1999).

Cut Scores

Additionally, the same principle regarding validity, reliability and fairness applies generally to the establishment and use of cut scores (lowest passing score) for the purpose of making high-stakes decisions. Cut scores, also known as cut points or cutoff scores, are specific points on the test or scale where results are used to divide levels of knowledge, skill, or ability. Cut scores are used in a variety of contexts, including decisions for placement purposes for other specific outcomes, such as graduation or promotion. Misclassification of students below the cut point can have negative consequences for students (*Designing school accountability systems: Toward a framework and process*, 2002). According to the U.S. Department of Education, Office of Civil Rights (2000), there is no single right answer to questions of when, where, and how cut scores should be set on a test with high-stakes consequences for students (*The use of tests when making high-stakes decisions for students: A resource guide*

for educators and policy makers, 2000). Thus, the inappropriate setting of cut scores on high-stakes graduation exit exams at some arbitrary point causes no further educational reforms and causes some students who should graduate from high schools not to do so.

Testing-Limited-English-Proficient Students

Students with diverse cultural and other background experiences, including variations in amount, type, and location (home country and United States) of formal elementary and secondary schooling, as well as interrupted and multi-location schooling of students (frequently experienced by children of migrant workers), are affected by language literacy, the contextual content of items, and the academic foundational knowledge base that can be assumed in appropriately interpreting the results of educational achievement tests. The format and procedures involved in testing can also affect accuracy in test scores, particularly if the test practices differ substantially from ongoing instructional practices in classrooms, including which accommodations are used in the classroom and how they are used. The U.S. Department of Education, Office of Civil Rights informs educators that the following factors are related to accurately testing limited English proficient students: language proficiency, cultural issues and schooling issues (*The use of tests when making high-stakes decisions for students: A resource guide for educators and policy makers*, 2000).

Researchers, educators, and parents remind us that the opportunity to learn does not begin in high school. Assessment standards, curriculum and instruction must be coherent and built logically on required knowledge and skills. They also insist that students should not be held responsible for information they have not been taught.

Texas and Accountability

The context or environment created by the Texas accountability system prior to 1984 was dominated by what former Commissioner of Education Lionel "Skip" Meno described as "the old way of doing things," that predicts students' success based on their parents, neighborhoods, and economic circumstances (*Equity driven achievement: Focus school districts*, 2000). The old way was a deficit model that assumed a bell curve for student achievement. Children of color from low-income homes were overwhelmingly on the low-performing end of the curve. And this was seen as a natural and inevitable consequence. In addition, the state's pre-1984 education system was almost exclusively based on process or input variables, dominated by efficiency (*Equity driven achievement: Focus school districts*, 2000). These, of course, are lower-level thinking skills.

The current Texas educational reform is rooted in two distinct conflicts. The first was the challenge of the unequal distribution of resources among Texas school districts. The state's largest minority groups—African-Americans and Hispanics—went to schools that received far fewer resources than schools that enrolled predominantly white students. The second conflict arose in the 1970s, when a group of new Texas businesses based on high technology and services challenged the hegemony of the state's traditional agricultural and oil interests (Carnoy, Loeb, & Smith, 2000).

In 1979, the Texas legislature passed the Equal Opportunities Act. However, the present educational reform began in 1984, with a push from H. Ross Perot, using his rationalist viewpoint, to bring Texas into the high-tech age and to resolve the pressures for equalized school funding by low-income groups (Carnoy, 2001). The reform went through two rounds (Reform Act of 1984 and

1987), and by 1991, was institutionalized in Texas politics. In addition to increasing funds for low-performing schools, the new learning standards for each grade were established, including statewide assessment and school accountability (Grissmer & Flanagan, 1998). Former Governor Ann Richards implemented this decentralization reform that gave the states control over standards and testing, but gave the schools choices in how to meet state goals (Grissmer & Flanagan, 1988).

Tests that are used to measure students' progress and hold educators accountable for raising student achievement play a key role in standard-based reform. However, according to Amrein and Berlin, the evidence found in their study of high-stakes testing in 18 states is that, in all but one analysis, student learning is indeterminate, remains at the same level as it was before the policy was implemented, or actually goes down when high-stakes testing polices are instituted. Clear evidence for increased student performance was not found (Amrein & Berliner, 2002). It is structurally misdirected because it treats the symptoms of school failure (i.e., inferior schools) rather than the cause (Valencia, 2001). This is particularly troublesome in Texas, because high school graduation rates are relatively low; according to Carnoy, Loeb and Smith, only about 65 percent of African-Americans and Hispanic eighth-grade students, and about 78 percent of white eighth-grade students, graduate four years later (Carnoy et al., 2001). A 35 percent dropout rate for African-Americans and Hispanics, as well as a 22 percent dropout rate for whites, should be unacceptable for all Texans. Nevertheless, we continue to use yearly averages of student test scores to identify low-performing schools and specify them for intervention.

On November 15, 2002, the Texas State Board of Education (SBOE) adopted passing standards for the Texas Assessment of Knowledge and Skills (TAKS) which provides for a two-year phase-in of the standards. Students who do not pass these exams will not receive a diploma. The projected impact of the new Texas assessment test (TAKS) is quite high. It is projected that, in grade 11, for Hispanics, 83 percent will fail math, 70 percent will fail Language Arts, 33 percent will fail Social Studies, and 67 percent will fail Science. African-American students, as well, are predicted to fare poorly on the new TAKS test; in grade eleven, it is projected that 80 percent will fail Math; 27 percent will fail Language Arts, 35 percent will fail Social Studies, and 75 percent will fail Science (State Board of Education, 2002).

Texas Testing Model

Texas has established a status accountability system, with high-stakes testing as its instrument of measure. According to the Council of Chief State School Officers (2002), there are four models for the accountability system. These models focus on status, improvements, student growth, or change of effectiveness. Model 1 (Status) examines how current students relate to the standard; Model 2 (Improvement) examines whether the school is getting better at helping successive groups of students meet the standards; Model 3 (Student Growth) measures whether students are learning from year to year; and Model 4 (Change of Effectiveness) examines whether students or subgroups make more than expected growth or rate of improvement increases. Texas's accountability system is built on Status, Model 1.

Brian Gong, Associate Director, Center for Assessment, during an interview at the Council of Chief School Administrators Conference in San

Antonio, Texas, (2003) stated that, during a peer review, he found that the Texas accountability system has a technical data-gathering system that is second to none, and the accountability system and assessment have improved. However, the state had several issues that must be addressed to ensure that all students have the ability to learn:

- Student dropout rate,
- Instruction,
- Support for the least advantaged students, and
- Definition of the academic year.

Theoretical Framework

Theoretically, the study is embedded in a political framework, voicing the need for a learning system within an educational democracy. The political frame views organizations as “alive and screaming” political arenas (Bolman & Deal, 1997). These arenas host a complex web of individual and group interests. However, this is the frame where the current dialogue and actions about high-stakes assessment currently resides. Business is demanding a more scientific management approach to monitor schools. Thus, the federal government and state legislative bodies have employed scientific management to resolve the nation’s educational dilemma: poor student performance and school accountability, using high-stakes testing, is the vehicle of choice. The rationale underlying high-stakes testing policies is rooted in the philosophical presumptions which link legislated high academic standards and expected high performance of all students. Educational policies’ efforts to mandate teaching and learning in state and federal programs have taken the form of a state accountability system, yearly program evaluation, and competency exams. The

public is assured of agencies' ability to monitor schools' performance by public reporting of high-stakes testing results, including ranking and/or grading schools.

Grassroots movements, some from the largest educational systems (e.g., Michigan, New York), find that the high-stakes testing movement is oppressive, and are calling for democracy in schools. They stress that high-stakes testing undermines the high quality of education and genuine student/teacher motivation. They warn that schooling has been simplified as nothing more than a chase for acceptable numbers, a dedicated means to exclude—to rank—the less powerful.

African-American, Hispanic and other ethnic groups have called for justice and equality. The Critical Race Theory provides a contextual understanding of contemporary legal debates concerning the effectiveness of past civil rights strategies in current political climates (Scheurich, 2002), including high-stakes testing. Minorities once again demand to know what is being done to close achievement gaps. African-Americans, Hispanics and other minorities question if all students are given the opportunity to learn the standards before they are tested. They challenge the extent to which testing meets the needs of all students, especially those who speak English as a second language and those existing in traditionally marginalized settings. They emphasize that certain consequences of high-stakes standardized testing promote a set of conditions that are unjust, unequal and conforming.

African Americans and Hispanics point out that legitimate learning necessarily presents itself in and on the basis of test scores, such testing refuses to admit and accept cultural and individual differences in knowledge, values, experiences, learning styles, economic resources, and access to the dominant

academic artifacts that ultimately contribute to both the appearance of achievement and the status of cultural hegemony (Vinson, Gibson, & Ross, 2001).

Both groups speak from experiences framed by racism, and come from a different frame of reference and voice than the dominant culture; and their complaints deserve to be heard. Some minorities agree with Gloria Ladson Billing's article, "Toward A Critical Race Theory of Education," and note that racism is endemic and deeply ingrained in American life; race continues to be a significant factor in determining inequity in the United States, and that class and gender-based explanations are not powerful enough to explain all of the differences or variances in school experience and performance (Ladson-Billings & Tate, 1997). In effect, standardized testing encourages a singular and homogenous public schooling—one antithetical to such contemporary ideas as multiculturalism, and differences and diversity underlying a technology of oppression.

In *Pedagogy of the Oppressed*, Paulo Freire referred to such standardization as a "banking system" in education.

> Here, schooling turns [students] into "containers," into "receptacles," to be "filled" by the teacher . . . The more completely the teacher fills the container, the better teacher he/she is. The more meekly the receptacles permit themselves to be filled, the better students they are . . . Education becomes an act of depositing, in which the students are the depositories and the teacher is the depositor . . . the scope of action allowed to the students extends only as far as receiving, filing and storing deposits (Freire, 1970).

Some economists are challenging whether or not Americans are receiving proper return on the monetary investment they have placed in education,

especially the education of minorities. Many question the monies allotted to Title I programs and programs like the “Robin Hood” plan (Chapter 41) in Texas.

Some educators feel high-stakes testing narrows the taught curriculum and is contradictory to what is best for children. They would prefer a learning system. A learning system is defined as the following (Bowsher, 2001):

A learning system integrates system designed principles and instructional design methods to produce a set of group learning sessions, individual learning models, interactive tutoring, and high quality motivational course materials. All of these elements are essential to enable all students to become successful learners by achieving learning objectives derived from educational standards.

A learning system is in the business of student learning, rather than just teaching.

Next, the study is embedded in a legal frame. Educational reformists cannot possibly hope to close the achievement gap without legal help. The current demand for accountability in schools is based on divergent values and belief systems that have determined educational politics in America for 150 years. These values have created stress upon the system, often making it difficult to answer the questions:

- Who should be taught?
- What should be the purpose of schooling?
- What children should be taught?
- Who decides issues of school policy and direction?
- Who pays for schools?

The answers to these questions are embedded in the school’s, community’s, or nation’s beliefs about choice, quality, efficiency and equity.

Educational conflict arises when beliefs about choice, quality, efficiency and equity are not aligned. Overt conflict between these beliefs is perhaps best represented by states' debates over high-stakes testing (Stout, Tallerico, Scribner, 1995). Over the last century, the courts could only resolve some conflicts.

According to Bolman and Deal, propositions summarize the political perspective (Bolman & Deal, 1997):

- Organizations are coalitions of various individuals and interest groups.
- There are enduring differences among coalition members in values, beliefs, information, interests, and perceptions of reality, the allocation of scarce resources—who gets what?
- Scarce resources and enduring differences give conflict a central role in organizational dynamics, and make power the most important resource.
- Goals and decisions emerge from bargaining, negotiation, and jockeying for position among different stakeholders.

The political frame asserts that, in the face of enduring differences and scarce resources, conflict is inevitable and power is a key resource (Bolman & Deal 1997). The question is: Who will receive the scarce educational resources provided by the federal, state, and local government? Will money be allotted to enable African-American, Hispanic and poor students to master the standards?

The final proposition of the political frame emphasizes that organizational goals are set not by the top but through an ongoing process of negotiation and interaction among key players (Bolman & Deal 1997). Hence, we see negotiation

between the federal government, states, and local school districts. Usually, these negotiations revolve around money; hence, the need to look at high-stakes testing and resources.

Changing to a learning system requires changing paradigms. According to Joel Baker, “A vision without action is merely a dream. Action without a vision just passes time. Vision with action can change the world.” (Bowsher, 2001, p. 259). It is not necessary to change the world, but there is a critical need to bring systemic change to the American public school system. Fulfilling this vision requires vision and realistic plans based on change management principles. Thus, it is important to study the unintended consequences of high-stakes testing on students, specifically African-Americans, Hispanics and the poor.

Summary

The review of the research reveals debates and studies, embedded in a political framework, focused on national and state high-stakes testing policies, with power groups for and against high-stakes testing in high school. In a response to demands from communities and businesses to make schools more accountable, many have adopted high-stakes testing policies.

The review of the research begins with a description of milestones in assessment-based reform from the 1950s to the present. It identifies recurring patterns of the states and schools, and their dependency on the use of tests and test scores to make judgments regarding students’ abilities, program placements and high school graduation (*GI Forum v Texas Education Agency*, 2000; Grissmer et al., 2000; Haney, 2000; Klein, 2000). The literature review revealed that many studies have concluded that high-stakes testing may have significant negative consequences for some students (Amrein & Berliner, 2002; Carnoy et

al., 2000, Carnoy & Smith, 2001; Orfield & Krohaber, 2001; Sheldon & Biddle, 1998). These consequences of high-stakes testing include:

- narrowing curriculum and changing instructional practices to address the test (Corbett & Wilson, 1990, 1991);
- reduction of time available for ordinary instruction, (Shepard, 1992; Smith, 1991), causing many to fear that students will be harmed by high-stakes testing (e.g., students will be retained in grades or denied diplomas because their school did not expose them to the skills and learning needed to pass the test) (Heubert, 2001);
- increasing the dropout rate (Kreitzer et al., 1989); research also has linked failure on exit exams with increased dropout rates among high-achieving minority students;
- an increased quest for money to fund educational reforms to ensure that all students have the opportunity to learn, and dispelling the myth that money does not matter (Hedges & Greenwald, 1996; Hedges & Greenwald, 1992) in the implementation of high-stakes testing (Heubert & Hauser, 1999).
- a majority of studies have found no evidence that exit exams increase student learning, as measured by other indicators such as standardized tests (Jacob, 2001; Neill & Gayler, 1999). One study examined the effects of high-stakes tests on learning by examining indicators of the students' learning, academic accomplishment, and achievement other than the tests associated with high-stakes, such as ACT, SAT, NAEP and AP exams; this study found 67 percent of the states that use high school graduation exams posted decreases

in ACT performance, and 56 percent of the states that use high-stakes high school graduation exams posted a decrease in SAT performance after those exams were implemented;

- NAEP mathematics data indicated that high-stakes testing policies did not generally improve the performance of students on math or reading tests;
- percentages of students passing various AP exams do not indicate that high-stakes high school graduation exams improve student achievement (Amrein & Berliner, 2002);
- exclusion rates of LEP students and students with IEPs (i.e., eliminating large numbers of Hispanics and African-American students) explained meager gains in the NAEP scores (Haney, 2000).

High-stakes testing in accountability systems is taking place at a rapid rate within a complex context, illustrating political, legal, cultural, and other contextual circumstances. It is linked to financial lawsuits; a drive to increase what students can do; a preface to taking strong action; a means for addressing inequities between schools or subpopulations, existing legal statutes, or regulations; and/or a way to validate a generally strong education system and challenge it to improve its capacities. The unintended consequences and outcomes reflect more than high-stakes testing. They are symptoms of deeper educational reform issues that must be addressed at the teacher level. Although there is research on high-stakes testing, studies examining the unintended consequences of high-stakes testing on poor and minority urban high school students from teachers' and students' points of view are missing. Thus, this study stands to make significant

contributions to educational research. More important is the insight it may provide schools regarding high-stakes testing and student success for African-American, Hispanic, and poor secondary high school students. Educational reform offers a real opportunity to improve student achievement. Each party involved in educational reform has much to offer. Approximately half of all states have “academic bankruptcy” regulations, allowing for state intervention (Fetterman & Smith, 2002). No longer are excuses for low-performing schools being accepted by the nation and states. States and districts are implementing educational reforms to meet legal requirements and public expectations. However, the implementations of these plans are often costly to African-American and Hispanic students, resulting in dropouts and lost educational opportunities. Just as the earning power of high school graduates has declined relative to that of college graduates (Hauser, 1993), so has the earning power of high school dropouts. The potential for the educational community is real.

CHAPTER III – METHODOLOGY

Introduction and Purpose of the Chapter

The investigation of the consequences of high-stakes testing (assessment) for high school students, specifically the unintended side effects of high-stakes testing for economically disadvantaged and ethnic minority students from critical perspective of teachers and students, is imperative and essential to successful education reform in America. It is important to study intended and unintended consequences of high-stakes testing on poor and minority student outcomes for the following reasons:

- high-stakes testing policy is regarded as a legislated instrument to improve education by making schools more accountable for the performance of their students;
- the possibility exists that high-stakes testing policy may be a symbolic response to very real educational problems (Ellwein et al., 1988);
- the need exists to examine how policy made at the macro-political (state and federal) level affects students at the micro-political level;
- legislated policy responses to educational problems often occur without sufficient study of the efficacy of the policy, its effects, or appropriateness to identify the problem (Ellwein et al., 1988).

This chapter describes the purpose, research paradigm, study design, and methodology used to explore complex interactions, processes and resulting student outcomes related to the implementation of a high-stakes testing policy at the secondary level.

Purpose of the Study

The purpose of this study is to determine the consequences of high-stakes testing on high school students, specifically those unintended side effects of high-stakes testing for economically disadvantaged and ethnic minority students, from the critical perspective of teachers and students. This research is organized around three related questions designed to target these conditions and their interrelationships. They are:

- What critical perspectives do teachers (and students) have about the effects of high-stakes testing?
- How do intended and unintended consequences of high-stakes testing affect students?
- How do intended and unintended consequences of high-stakes testing affect instruction and classroom practices?

Data collected on this societal trend and differential educational opportunity will be useful in America's educational reform process, especially in schools embarked on fundamental reforms in academic outcomes for students and committed to reducing and/or eliminating the achievement gap between whites and minority subgroups. Several studies have been conducted on high-stakes testing; however, these studies often do not include voices of students and teachers, essential elements in this area of educational reform. Information collected by the study will allow teachers to more readily equip students to become proficient in state standards. Additionally, it will also aid schools and faculty in choosing the appropriate pedagogy when instructing poor and minority children.

Framework

Qualitative research was chosen, for it enables the researcher to study social and cultural phenomena. Qualitative research methods are designed to help researchers understand people and the social and cultural contexts within which they live. Kaplan and Maxwell (1994) argue that the goal of understanding a phenomenon from the point of view of the participants and its particular social and institutional context is largely lost when textual data are quantified. A policy research framework was selected for the study because it allows for the examination of multiple dimensions of a policy. Policy research is defined as the "process of conducting research on, and analysis of, a fundamental social problem in order to provide policymakers with pragmatic, action-oriented recommendations for alleviating the problem."

Critical Paradigm

A critical paradigm was selected because, as noted by Oliver in Mertens (1998), it directly addresses the politics in research by confronting social oppression at whatever level it occurs. Critical or emancipatory researchers argue that the interpretive/constructivist paradigm did change the rules; however, it did change the nature of the game (Mertens, 1998). As Foster (1993) stated, the paradigm goes beyond the issue of the powerful sharing with the powerless and relinquishing control of the research to the marginalized group (Mertens, 1998). They ask, "What is just? What can we do to collectively change the world?" Thus calling ideology into question, and initiate action, in the cause of social justice. This form of critical inquiry keeps the spotlight on power relationships within society so as to expose the forces of hegemony and injustice.

Human beings engaged in intervention in the world as transformers of that world-that results in the development of critical consciousness (Freire 1972, p.47).

Ontologically, reality is constructed and multiple realities are recognized. The paradigm reflects historical realism, shaped by the influences of social, political, gender, ethnic and disability values. The epistemology of knowledge is value-mediated and value-dependent. Minorities wish to understand the world view and assist in changing it. Their methodology raises questions to heighten awareness of injustice, and begins the change process.

Historically, minorities and the poor have been disenfranchised. Implementation of high-stakes testing as a graduation requirement has the potential for unintended consequences, many of which may be negative. Critical researchers assume that social reality is historically constituted, and that it is produced and reproduced by people. Although people can consciously act to change their social and economic circumstances, critical researchers recognize that their ability to do so is constrained by various forms of social, cultural, and political domination. The main task of critical research is seen as being one of social critique, whereby the restrictive and alienating conditions of the status quo are brought to light. Critical research focuses on the oppositions, conflicts, and contradictions in contemporary society, and seeks to be emancipatory, i.e., to eliminate the causes of alienation and domination.

Research Design Rationale

Case study research excels at bringing an understanding of a complex issue or object, and can extend experience or add strength to what is already known through previous research. Researchers have used case studies to examine contemporary real-life situations and provide the basis for the

application of ideas and extension of methods. Researcher Robert K. Yin defines the case study research method as an empirical inquiry that investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used (Yin, 1984). According to Merriam (1988), case study research in education seeks to understand specific issues and problems of practice.

Well-known case study researchers such as Robert E. Stake (Stake, 1995), Helen Simons (Simons, 1980), and Robert K. Yin (Yin, 1984), have written about case study research and suggested techniques for organizing and conducting the research successfully. Yin's definition obviously specifies that case studies are appropriate research tools when the researcher deliberately wants to address contextual conditions considered important to the phenomenon being studied. To investigate the object of the case study in depth, researchers should use a variety of data-gathering methods to produce evidence that leads to understanding of the case and answers the research questions.

According to Stake, Simons, and Yin (1995, 1980, 1984), a key strength of the case study method involves using multiple sources and techniques in the data-gathering process. Data gathered is normally largely qualitative, but it may also be quantitative. Since this view and Yin's definition recognize the need to use multiple sources of evidence in order to develop a more holistic portrayal of events, a combination of both qualitative and quantitative evidence was collected. Tools to collect data included surveys, interviews, documentation review, observation, and even the collection of physical artifacts.

Finally, researchers from many disciplines use the case study method to build upon theory, to produce new theory, to dispute or challenge theory, to explain a situation, to provide a basis to apply solutions to situations, to explore, or to describe an object or phenomenon. The advantages of the case study method are its applicability to real-life, contemporary, human situations and its public accessibility through written reports. Case study results relate directly to the common reader's everyday experience, and facilitate an understanding of complex real-life situations.

Study Presuppositions

The following presuppositions guided the research study: (1) large-scale assessments may have unintended negative consequences for minority and poor students; (2) high-stakes testing policy does not necessary result in higher student achievement and performance for all students; (3) high-stakes testing policy does not necessarily result in educational benefits for all students; and (4) the response to high-stakes testing by central Texas urban schools is influenced by historical events and sociopolitical events which occur in community and state.

Selection Process

Selection of Site

Data from the Texas Education Agency was used to randomly select a site in Texas. The school must have a diverse student population of which large percent are minorities and/or economically disadvantaged students (students eligible for free or reduced lunch). The district should have an overall

“acceptable” accountability rating from the Texas Education Agency for its TAAS scores and dropout rate.

Sample Characteristics

Experienced high school teachers were chosen by use of purposive sampling selection process to be a part of the study. Both groups (students and teachers) have first hand experience with the effects of high-stakes testing on high school students. They will be able to provide information that can be used with wider populations. The sample is also strategic because those teachers who had less than one year of teaching experience were eliminated from the study; they do not have first-hand knowledge of high-stakes testing and its effects on students.

Interactive Qualitative Analysis (IQA): Overview

According to Yin, the amorphous nature or “softness” of the data has plagued the use of case study methodology and the debate regarding validity, and has been ongoing (Yin, 1984). Interactive Qualitative Analysis (IQA) was chosen as the data collection and data analysis tool because it has a unique innovative research design that avoids the pitfalls described by Yin and allows researchers to perform a more in-depth assessment of studied phenomenon. The researcher is allowed to look at a problem from the viewpoint of constituencies, power and distance, issues, comparisons, and research questions. Unlike many other qualitative research methods, it is not a linear, get-it-right-the-first-time process; it is more circular in nature, with recursive (looping) features that allow for successive project refinement (Northcutt & McCoy 2002). IQA provides a research process that is capable of transforming vague outcomes

often found in linear research methods into more and more precise, relevant outcomes. According to Northcutt & McCoy (2002), it allows successive refinement of each of the following:

- Constituencies
- Classification of Constituencies
- Issues
- Comparisons and Research Questions

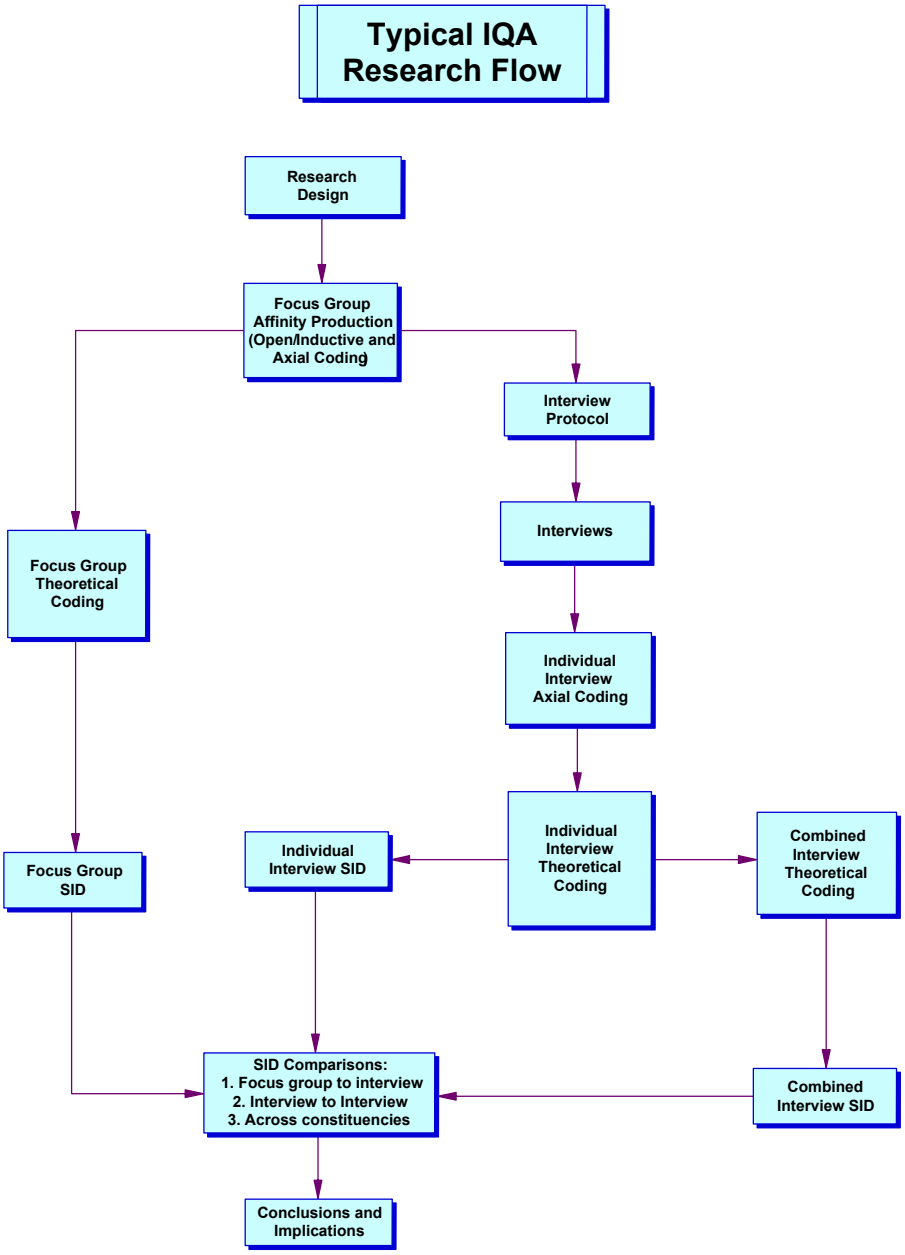
First, c*onstituency* reflects both interest and power over the phenomenon, which is the center of the problem statement (Northcutt & McCoy, 2002). Next, *sorting,* or classification of the constituencies in terms of both distance from and power within the phenomenon, is a useful design exercise. Distance and power analysis helps to ensure consistency among the purpose of the study, field study methods, and analytical procedures (Northcutt & McCoy 2002). Simply put, the study investigates the purpose the researcher wants it to address in a systematic way. Thirdly, according to Northcutt and McCoy, different constituencies have different perspectives on the same phenomenon, so the issue statement must be meaningful to each constituency. The *issue* statement is always a version of "Tell me about . . ." but it must be presented in terms that are real to a given constituency. Lastly, *comparisons* will generate research questions. Three general research questions are: (1) What are the components of the phenomenon? (2) How do the components relate to each other as a system? And if more than one consistency made up a system, (3) How do the systems compare in terms of parts—intra-systemic relationships and inter-systemic relationships (Northcutt & McCoy 2002)?

IQA is a system that has clearly defined elements. It uses a “mind-map” to allow the reader to conduct a “tour” through the system in which relationships of each element can be viewed and analyzed. The IQA System allows comparisons to be made on two levels: (1) comparing individual mind-maps to each other and the composite mind-map (qualitative analogue to the statistical variations); and (2) comparing composite mind-maps of different constituencies (Northcutt & McCoy, 2002). The IQA system allows the comparison of individual and group realities.

The IQA system has several components. The study usually begins with a focus group of people who share some common experience, work or live within some common structure, or have a common background. Two major outcomes are generated in the research process: (1) identification of factors or affinities; and (2) identification of relationships among factors (Northcutt & McCoy, 2002).

Below is an example of an IQA system:

Figure 3.1
Typical IQA Research Flow



The IQA system allows the identification of system elements, describing the relationships, and understanding how the elements and relationships interact within a system. It provides a reliable, dependable, transferable way of

representing realities of individuals or groups while highlighting the concepts of validity and reliability. Thus, IQA was chosen as the method to collect and analyze data in this study because of the above-mentioned characteristics, and because of its circular, recursive (looping) features that allow for successive project refinement. Additionally, IQA provides a research process that is capable of transforming vague outcomes into more precise, relevant outcomes.

The Researcher as Instrument

In qualitative studies, the researcher is the "instrument": my presence in the lives of the informants invited to be a part of the study is fundamental to the paradigm. In this study, several important points regarding the researcher as the instrument need to be stressed. First, I am African-American, born and reared in the south, and experienced in teaching culturally diverse students. Additionally, for the last five years, I have been district curriculum specialist, supervising and assisting teachers with K-12 curriculum and instruction for more than 45 schools focusing on low-performing schools. These factors make me an insider.

There are some advantages to being an insider while conducting field research (Delgado-Gaitian, 1993), especially in a minority community. My familiarity with and knowledge of the culture, having been reared as an African-American, have also been advantageous to formulating questions during the interview.

When any researcher enters into the lives of the informants, there is a range of strategic, ethical, and personal issues that does not lend itself to qualitative approaches (Locke, Spirduso, & Silverman, 1993). These issues then can be sorted into technical ones that address entry and efficiency in terms of roles, and interpersonal ones that capture the ethical and personal dilemmas that

arise during the conduction of the study, as suggested by Marshall and Rossman (Marshall & Rossman, 1994). As an insider in the community and the district being studied, the researcher needed to fewer challenges negotiating with the schools and overall organization. A living history of the district enabled the researcher to better determine the varying degrees of actual participation in the daily lives of the participating teachers. As an insider, there are also visible advantages to maintaining good relationships, respecting norms of reciprocity, and sensitively considering ethical and political issues.

Some of the disadvantages of being an insider, however, also presented challenges. Although a district and ethnic insider, I was still an outsider to certain specific communities, organizations, and individuals. Some teachers were reluctant to talk about past administrators, narrowing of curriculum, and other sensitive issues that emerged during the study. Some individuals were reluctant to provide permission to record conversations, and others tended to express distrust regarding the study. The political dimension of the research emerged as an issue for some informants. Who is going to use the data, for what purpose, and is this a ploy to remove teachers the district considered ineffective? Therefore, subjectivity and rapport were issues to be considered.

Study Procedures

This study involves an IQA system, which allows the comparison of individual and group realities of teachers and students of intended, as well as unintended or negative consequences of high-stakes testing. IQA data collection techniques originated from Total Quality Management (TQM). TQM processes are designed to capture knowledge from organizational members to solve problems and improve processes. The basic assumption of TQM is that people

who are closest to the problem best understand how to fix it. Informants are allowed to describe and label their experiences and identify relationships between their experiences and produce a conceptual map. Categories of meaning are called *affinities*.

There were three focus groups in the study—teacher and student—residing in a large urban Texas school district. It is hoped that the results of the study will allow researchers and other educators to identify the negative and unintended consequences of high-stakes testing, as well as possible strategies to foster educational reform.

Step 1 – Initial Stages of Entry and Access

Preliminary meetings with the superintendent, principal and other school administrators were held to obtain permission to conduct the study within the district. Selection criteria for teachers in the study were discussed and agreed upon. The first focus group consisted of eight high school students who had failed the Texas graduation exam. The second focus group consisted of ten high school students who had passed the Texas graduation assessment. The third focus group in the study consisted of ten teachers purposely selected from the district school population. They met the following criteria: (1) certified high school teacher; (2) minimum of two years' teaching experience; (3) experience working with diverse school populations; and (4) familiarity with the Texas accountability system. Throughout the study, communication among administrators, central office personnel, and myself was maintained via personal conferences and/or visits, e-mail, telephone calls, and correspondence.

Step 2 – Calendar

A calendar for the study was designed. The calendar was approved by school personnel and the University of Texas dissertation chairman.

Step 3 – Study Informants

Study informants were selected using a purposeful selection process for an southeast Texas school district. Letters were sent to each informant describing the study and time required, and asking that they become a part of this vital research.

In addition, a site for the focus group meeting was selected and ultimately utilized.

After confirmation was received from teachers and students who agreed to participate in the study, thank-you notes and notices of the time and place of focus group meetings were sent.

One week before the focus group meeting, all informants are sent reminder notices of the focus group meeting via e-mail or telephone.

Step 4 – Pilot Study

A pilot study was be conducted to refine the focus group issue statements and to adjust aspects of the study where needed, using IQA methodology. Purposeful selection was used to select ten informants from the central Texas teachers' association and ten students from a central Texas urban high school. For the focus groups, facilitated discussion was used. The teacher discussion began with the following issue statement:

Legislators and policymakers have expressed the intended goals for high-stakes testing, as improving student learning and achievement, and making

schools and teachers accountable. However, they are often vague about implementation of such policies and sometimes do not provide adequate resources to help students close academic achievement gaps. Often, students arrive at high school without adequate academic skills. As teachers help prepare all students, including African-American, Hispanic or the poor, to be successful on high school high-stakes exams such as the TAKS test, what factors or challenges do they and their students encounter or experience that would be considered as negative consequences of high-stakes testing policies in America's effort to reform schools?

Students received the following issue statement:

"Please tell about your experiences with the State of Texas graduation assessment."

Sample interviews were conducted with both groups, and issue statements were refined.

Step 5 – Study Focus Group

IQA methodology was used to identify affinities. This included silent brainstorming; clarification of meaning phase; affinity grouping and affinity naming and revision were performed during the focus group. For the focus groups, facilitated discussion was used. Teachers (10) and students (20) comprised different groups, and the focus group study process occurred at different times and places.

First, the groups participated in *silent brainstorming*. Members of the groups were read the appropriate issue statement found in Step 4, and were asked to write down their ideas or experiences about the unintended or negative consequences of high-stakes testing on an index card (one experience per card).

During this period, individuals silently “brainstormed” by writing individual thoughts and reflections on index cards (silence reduces undue pressure by peers or facilitator). The group was given approximately ten minutes to record these thoughts in single words, short phrases, or even diagrams that came to mind regarding the issue statement. Focus group members were encouraged to produce as many cards as they wished; these were placed in a comfortable and safe environment, easing the task of data production and participatory research. The informants were asked to refrain from censoring their thoughts or responses. Examples of suggestions included, “All thoughts you have about high-stakes testing are okay.” **“**Don’t analyze; just write.” “No one will criticize your thinking.”

The facilitator monitored the group, and the process was concluded when it appeared that the informants had an opportunity to generate a *satisfactory* number of responses.

When informants had stopped writing, it was time to proceed to the next step. Research has found that the group process is effective, for it encourages the maximum production of individual thoughts, feelings and ideas, yet creates a coherent group construction disparate with these individual realities (Northcutt, 2002).

Next, *the clarification of meaning phase* was begun by the facilitator, who read a card and allowed the group to come to a consensus as to the meaning of each card. The focus group members were instructed to tape the index cards on the wall in rows and columns so that participants could view them. Through a group discussion, the facilitator guided informants in clarifying their understanding of the responses on each card in order to eliminate any ambiguity and vagueness associated with the meanings of the words or phrases. The

researcher then facilitated a data clarification process. According to Northcutt, this process offers two advantages (Northcutt & McCoy, 2002):

The discussion creates a more full-bodied understanding of the written comments and a shared understanding of the responses.

The discussion engages participants who need to interact to gather their thoughts, and prompts additional data to be produced.

This procedure was implemented in this study. Even though each card was written by an individual, the anonymous author had no more claim to the meaning of the card than any other group member did; therefore, anyone in the group could offer an opinion about the meaning of a particular card.

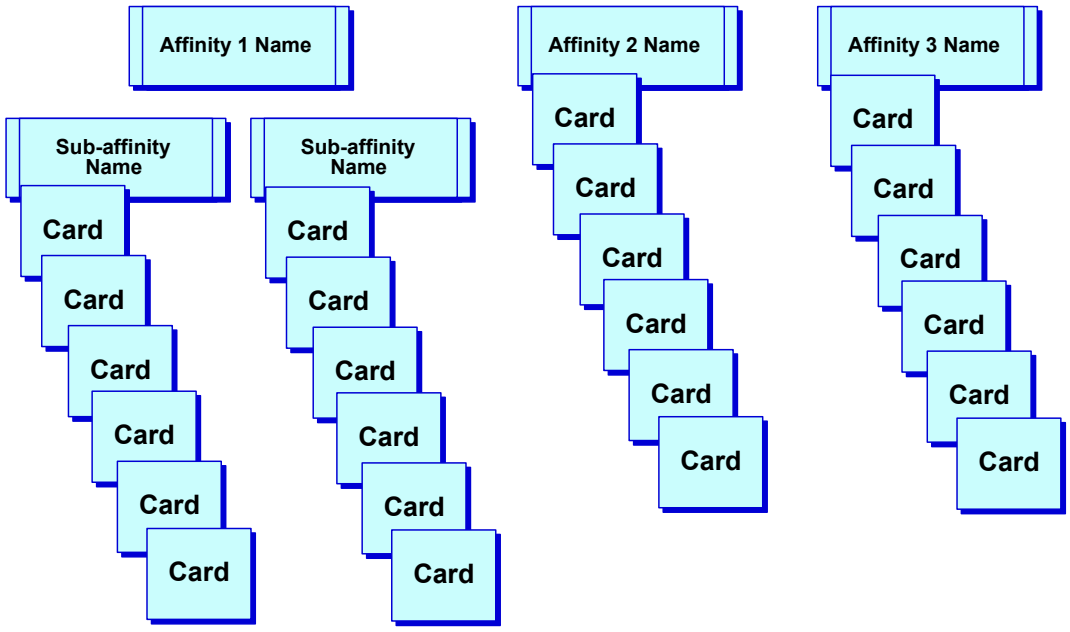
After the clarification conversation and/or process, the participants were able to choose to add more reflections and thoughts to the original body of index cards. The facilitator encouraged any further production of responses and a second clarification discussion, as necessary, to ensure that the responses reflected the individual and shared experiences of the group members relative to the issue statement.

The third function of the focus group was affinity group or inductive coding. The group will be asked to silently organize the cards in groups by meaning. *Coding* is the name given by qualitative researchers to describe the way in which text is represented by abstractions (Northcutt & McCoy, 2002). Coding demands both induction and deduction. The cards represent isolated, fragmentary facts. They are data, the raw material of reflection; their lack of coherence perplexes and stimulates further reflection. Their possible meaning suggests a mental platform, an intellectual point of view, from which to note and define the data more carefully, to seek additional observations, and to institute, experimentally,

changed conditions. There is a double movement in all reflection: movement from the given partial and confused data to a suggested comprehensive (or inclusive) entire situation; and movement back from this suggested whole—which, as suggested, is meaning an idea—to the particular facts, so as to connect these with one another and with additional facts to which the suggestion has directed attention. Roughly speaking, the first of these movements is inductive; the second, deductive (Northcutt & McCoy, 2002).

Finally, grouping is followed by the affinity naming and revision phase (axial coding), which consists of giving a name to the group (affinity) and sorting any cards that may have been placed in the wrong group. An example is shown in the figure below.

Figure 3.2
Affinity Cards



Source: Northcutt and McCoy, 2002

A well-identified affinity has several characteristics (Northcutt & McCoy, 2002):

- It is **not** a person, a place, or a physical thing (except, perhaps, metaphorically); rather, an affinity describes constructs or characteristics of categories of meaning.
- It is homogeneous—it is about one construct rather than a mixture of topics.
- It is easy to define. If it is difficult to name or point to several different things, most likely, it is a mixture.
- On the other hand, it should have a range of meaning within this definition. For example, rather than allowing two affinities such as *Positive Emotions* and *Negative Emotions,* one affinity—*Emotions*—suffices.
- It has context (relationship) to other things; but affinity descriptions should not include theoretical codes within axial ones.

In this research, the entire focus session for each group lasted approximately 1.5–2 hours.

Step 6 – Pair Relationships

With the affinity groups defined, the next phase was the identification of relationships between each of the affinities. The rules for analyzing all possible pairs guided this phase:

- A ⟶ B (A influences B)
- A ⟵ B (B influence A)
- A <> B (No relationship)

An arrow may not go both ways. Sometimes, based on information received from the subject, it may not be obvious in what direction the arrow should be placed (Northcutt & McCoy, 2002). In this study, when this occurred, a separate list will was kept and a secondary interview was conducted.

Step 7 – Constructing the Interrelationship Diagram (IRD).

An Interrelationship Diagram, or IRD, is a table that represents all the relationships among the affinities. Creating an IRD is the first step in rationalizing the system process. The output of the focus group's hypothesizing activity is summarized in an IRD, which depicts the relationship or lack of relationship between affinities.

In this study, each focus group investigated links between the affinities by developing propositions (statements of cause and effect) from their own data. This activity, called theoretical coding, was employed: to create an extended reality for the group through further discourse. Using a forced directional choice in a specific order, the researcher was able to evaluate whether there was a direct cause/effect relationship or whether no relationship existed. According to Northcutt's description, the goal was to identify the skeleton of a "theory in perception" (Northcutt & McCoy, 2002). Theoretical coding of the affinities resulted in an Interrelationship Diagram.

The arrows show whether each affinity in a pair is perceived as cause or effect, or if there is no relationship between the pair. Placing arrows in a table showing the direction of relationships creates the IRD.

An arrow pointing from A to B (A → B) indicates that A is a cause or influencing affinity, and that B is the effect or influenced affinity. There are only

two ways arrows may point on the table: **left** or **up.** Next, each relationship is recorded twice (double bookkeeping). For Example:

Table 3.1
IRD Table Part-1
Tabular IRD

	1	2	3	4	5	6	7	8	9	OUT	IN	Δ
1		←										
2	↑											
3												

The sample table indicates that there is a relationship between 1 and 2. It reads that 2 influences 1. Please note that the arrows are pointing away from the 2 and toward the 1. Each relationship is identified and placed in the table.

Table 3.2
IRD Table Part-2
Tabular IRD

	1	2	3	4	5	6	7	8	9	OUT	IN	Δ
1		←	←	↑	↑	<>	←	<>	<>			
2	↑		←	↑	←	←	←	←	<>			
3	↑	↑		↑	↑	<>	↑	<>	←			
4	←	←	←		↑	←	←	↑	↑			

Next, the arrows are counted to find the delta value. The delta value is calculated by following the rules below:

- Count the number of up arrows (↑) or Outs
- Count the number of left arrows (←) or Ins
- Subtract the number of Ins from the Outs
- to determine the (Δ) deltas
- Δ = Outs – Ins

Table 3.3 IRD Table Part-3 Tabular IRD												
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**OUT**	**IN**	**Δ**
1		←	←	↑	↑	<>	←	<>	<>	2	3	-1
2	↑		←	↑	←	←	←	←	<>	2	5	-3
3	↑	↑		↑	↑	<>	↑	<>	←	5	1	4
4	←	←	←		↑	←	←	↑	↑	3	5	-2

Next, the table is then sorted in descending order of the delta.

Table 3.4 IRD Table Part-4 Tabular IRD												
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**OUT**	**IN**	**Δ**
3	↑	↑		↑	↑	<>	↑	<>	←	5	1	4
1		←	←	↑	↑	<>	←	<>	<>	2	3	-1
4	←	←	←		↑	←	←	↑	↑	3	5	-2
2	↑		←	↑	←	←	←	←	<>	2	5	-3

It is an easy and logical way of keeping track of relationships between affinities, and a great way to document qualitative research.

Step 8 – The Interview

Constructing an Interview Protocol

In Step 8, interviews were conducted with each participant in the focus group. These interviews are critical components of this qualitative study. The content of the interview is determined by the affinities developed by the focus group. The definition of each affinity was shared with each participant. The researcher was able to gather additional information by asking open-ended question like, “Tell me more about it,” or, “Tell me, what do you mean by this?”

The interview was a semi-structured interview, and it was designed to capitalize on the consistency afforded by a highly structured interview and the level of detail offered by open-ended or emergent interviews. The interview questions were designed and based on the affinities and sub-affinities developed by the focus group members. The interview protocol for this study was designed to achieve specific objectives, each of which related directly to the research questions of the study. In particular, interviews served to:

- Add richness and depth to the description of the meaning of affinities that was not possible with a focus group alone;
- Allow for individual mind-maps, which could be used in a debriefing session as an interpretive aid to the investigator.

The affinities produced by the focus group were used to create an interview protocol. The interview protocol was used to confirm affinities created by the focus group and to elicit descriptions of relationships among the affinities. The purpose of the protocol was to use the affinities identified through focus group data collection and analysis to inform and shape questions for the second round of data gathering: the interview. The focus group data collection served as research for the interview as well as a pilot study, providing a snapshot of the group mind-map.

The interview protocol should consist of two parts (Northcutt & McCoy, 2002):

- Axial Interview—an open-ended interview design to provide rich description of affinities by the respondents; and
- Theoretical interview constructed to identify relationships between affinities.

This structure was followed in this study.

Conducting Interviews

Overview

The interview consisted of two kinds of questions: What does this affinity mean to the subject (Axial code question), and, How does this affinity relate to all others in the system of perceived cause and effect (Theoretical code questions)? Logistical and operational details are required for successful a interview. The researcher/interviewer must get the interviewee to construct his or her reality. From Northcutt, a checklist highlighting the most important logistical issues to consider in the interview (Northcutt & McCoy, 2002) is shown below:

- Establish a relaxed atmosphere that encourages the interviewee to respond to each question sincerely;
- Memorize or become familiar with the interview protocol, including affinity names, their descriptions and the general order they should be covered in the interview;
- Test your equipment; make sure it works, and works well;
- Introduce yourself and provide basic information regarding the project;
- Explain the confidentiality of interviewee's responses, and get permission to tape the interview;
- Get to know something about the respondent; do not forget to gather demographic information;
- Start with the primary driver of the protocol;
- Always wait for the interviewee to finish speaking before you probe or move on to the next question;

- Probe for deeper meaning or extended examples;
- Provide examples if the interviewee does not respond or is unclear;
- Make a pre-empting statement to give the respondent a few seconds to gather thoughts;
- Summarize when making a transition to the next topic; be sure to tell the interviewee what will be asked next;
- Be especially attentive to metaphoric language;
- Remember to include the sub-affinities;
- Thank the interviewee for his or her assistance.

Northcutt and McCoy (2002), remind the interviewer to avoid talking too much; asking yes and no questions; asking questions that may divert focus of the interview; failing to ask for examples, and failing to be attentive to interviewer fatigue. All of the guidelines were followed in this study.

Interview Procedures

For this study, the interview procedure included: preparing for the interview; opening the interview; conducting the axial interview; conducting the theoretical interview; performing the interview wrap-up; debriefing; and documentation of the interview. The researcher successfully conducted all these activities.

According to Northcutt, preparing for the interview involves two areas: content familiarity and logistics setup (Northcutt & McCoy, 2002). Content familiarity is critical to the success of the interview. The interviewer must obtain a thorough understanding of the interview protocol and each affinity within it. The logistics set-up includes performing a sound and equipment check for recording the interview, securing a new tape and batteries to eliminate potential problems,

and having a printout of the affinity relationship table. The researcher should open the interview by establishing rapport, expectations and agreements, and initiating:

- A friendly greeting, self-introduction, description of the research study and the respondent's role;
- Discussion of confidentiality, anonymity for the interviewee, use of a tape recorder to record the discussion, and a description of how interview results will and will not be used.

As Northcutt relates, once the interviewee agrees to the conditions, the axial interview will begin. The interviewer begins the interview with a question about the primary driver. For example, "Tell me about your experience with affinity XX (the driver)." The interviewer listens to the respondent's answer, and asks for clarifying or confirming statements. After the axial interview has been completed, a short break is taken. Next, the theoretical interview is conducted. The interview is begun by the interviewer giving the interviewee a copy of the Affinity Relationship Table to guide the second phase of the interview, which is the examination of perceived relationships between all possible pairs of affinities by the respondent under the guidance of the interviewer. Begin the theoretical coding phase with an introduction, such as:

> Now that we have talked about each affinity, I would like to explore the connections you see between them. For example, in our earlier discussion, you mention a connection between affinity A and affinity B. This suggests that you see a relationship between the two affinities. I would like you to work through these pairs with me, and tell me what you see as the connections.

Next, the study continues by proceeding through the Affinity Relationship Table (ART). The ART can be completed in real time by the researcher or an assistant. Next, the interview is wrapped up by the interviewer asking the

interviewee for any final thoughts, and thanking the respondent and reiterating the confidentiality guidelines.

The analysis of the interview proceeds exactly parallel to the manner of the focus group protocol. For each of the affinities, the interview respondent is asked three kinds of questions (Northcutt & McCoy, 2002):

- What does this mean to you?
- What led to this?
- What are the results?

In a manner analogous to the focus group's activities, the interview transcript is coded both axially and theoretically. All of these guidelines were followed in the conduct of this study.

Step 9 – Axial Coding

The Axial Code Table (ACT) is the primary documentation for all comments that illustrate the range of meaning of each affinity for each respondent. The researcher identifies axial codes by noting key words or phrases that describe or illustrate an affinity. This text is then documented in the Individual Interview Axial Code Table (ACT). Quotes relating to a specific affinity are placed into the ACT, along with the line(s) of the transcript that were the source of the axial quote. Next is a sample Axial Code table:

Table 3.5
Graduate Student Relationships
Interview #24
Axial Code Table

Affinity	Transcript Line	Axial Quotation	Researcher Notes
1. Caring	130-131 132-133 136-139	*Well, I'm a word person. You have to tell me you love me. I have to hear that. You have to tell me you re in my corner*. *I'm also a gift person. I bring things to people. That's my way of saying I care about you*. Caring to me is really important because it drives me to not make excuses. *People care about me so I don't have any excuses as to why it can't be done. It's one of the motivating forces for me in this program.*	***"People care about me so I don't have any excuses as to why it can't be done"*** The student describes caring by what is said to her. And she states that she demonstrates her caring for others by her gifts. I was surprised that she did not include the "actions" of people and of herself as a form of caring.

Step 10 – Theoretical Coding

The System Influence Diagram (SID) is a picture drawn using a set of rules for rationalization on a summary of the theoretical codes called an *Interrelationship Diagram (IRD)* produced by the focus group. *Theoretical coding* refers to documenting the perceived cause-and-effect relationships among all the affinities in a system. In the focus group setting, this is accomplished by

facilitating a systematic process of building hypotheses linking each possible pair of affinities (Northcutt & McCoy, 2002). The group *Interrelationship Diagram (IRD)* summarizes the results of group theoretical coding. IQA resolves three issues in respect to theoretical coding (Northcutt & McCoy, 2002):

- What level of detail is desired in constructing each perceived relationship?
- How will the group be organized for analysis of relationships?
- How will a group composite (the system that represents the entire group) be constructed?

All possible direct links between the affinities are investigated by developing hypotheses grounded in the data. The relationships among affinities are analyzed using the "If . . . then . . ." or hypothetical construction. Hypotheses are recorded on a protocol called the *Affinity Relationship Table (ART)*. IQA provides focus group informants with a formal protocol to determine whether or not there is a direct influence between every possible pair of affinities in the system. The focus group then determines the directionality of influence.

A "detail" ART is produced for each focus group. In the detail ART, each focus group member is asked to determine the nature of the relationship between all possible pairs of affinities as in the previous form. It requires more time than the simple one because informants are asked to write a statement—preferably in hypotheses form for consistency and clarity of logic's sake—that reflects their experiences which support the cause-and-effect relationship recorded.

The Theoretical Code Table (TCT) is the primary documentation that illustrates the manner in which the affinities are related for each respondent. Here is an example of a TCT:

Table 3.6 Theoretical Code Table Graduate Student Relationships Interview #24 Theoretical Code Affinity Relationship Table			
Affinity Pair Relationship	**Line Number**	**Theoretical Quotation**	**Researcher Notes**
1 → 5	264	*Caring affects roadblocks because roadblocks shrink smaller when there is care.*	***". . . roadblocks shrink smaller when there is care"*** I do not know if roadblocks shrink and become smaller when someone cares about them, but their caring makes the roadblock easier to bear.
2 → 4	279	*Emotions affect perseverance*.	Negative emotions will have a negative affect on the student's resolve.
2 ← 5	281 283 - 287	*Roadblocks affect my emotions* *When people purposely put something . . . you know you have to do a 10-page report on your job and no one else has to do a report, I get angry and anger is not a good thing for me because I'm not productive when I am angry*. So, roadblocks trigger things in me that I don't like and they cause me to do mean and evil things like take stuff down in my office.	The statement, **"So roadblocks trigger things in me that I don't like and they cause me to do mean and evil things like take stuff down in my office**," indicates that the student will allow her emotions to cause her to commit acts that may be damaging to her career.

During the second phase of the IQA interview: *theoretical codes*, which *illustrate a relationship between two or more affinities, t*he relationship reported by the respondent (using the same rules as the focus group) is recorded by placing the appropriate arrow in the Individual Interview Theoretical Code Affinity Relationship Table. This documents both the direction of the relationship and the example or line of reasoning given by the respondent.

Additionally, the interview transcript should be examined for statements that illustrate a link between affinities.

Step 11 – Constructing the SID

Next, the System Influence Diagram (SID), or mind-map, will be created. A SID is a visual representation of an entire system of influences and outcomes. The graphic representation of relationships paints a vivid picture of system dynamics for both investigator and informants, and lends itself readily to analyzing how modifications might change the nature of the system (Northcutt & McCoy, 2002).

The SID is roughly analogous to a set of *qualitative structural equations* or as a *path diagram*; however, it is distinguished from traditional path diagrams in that recursion or feedback loops are allowed (Northcutt & McCoy, 2002). The SID is a visual representation of the “theory in perception,” grounded in the specific experiences and logic of the informants.

A SID is made or created by taking the information found in the IRD and presenting it to a system representing affinities and relationships among affinities. The SID is the central visual component in the IQA System. Affinities are identified as: Primary Drivers, Secondary Drivers, Secondary Outcomes, and Primary Outcome, based on their delta value. The value of delta is used as a

marker for the relative position of an affinity within the system. Affinities with a positive delta are *relative drivers* or causes; those with negative deltas and *relative effects* are outcomes.

According to Northcutt and McCoy (1998), the Tentative SID Assignments Table represents the initial placement of affinities for the SID. An affinity marked by a high positive delta or number resulting from many *Outs* but no *Ins* is a *Primary Driver:* a significant cause that affects many other affinities, but is not affected by others. *No Ins Rule:* Any affinity with no *Ins* is always a Primary Driver. The *Secondary Driver* is a relative cause or influence on affinities in the system; it is identified when there are both *Outs* and *Ins*, and there are more *Outs* than *Ins. Circulators/Pivots* occur when there are equal numbers of *Ins* and *Outs.* The *Secondary Outcome* reveals a *Relative Effect*; it is identified when there are both *Ins* and *Outs*, but there are more *Ins* than *Outs*. An affinity marked by a high negative number that results from many *Ins* but no *Outs* is a *Primary Outcome:* a significant effect that is caused by many of the affinities, but does not affect others. *No Outs Rule***:** Any affinity with no *Outs* is always a Primary Outcome. Since it has the highest negative delta, it is labeled *primary outcome*.

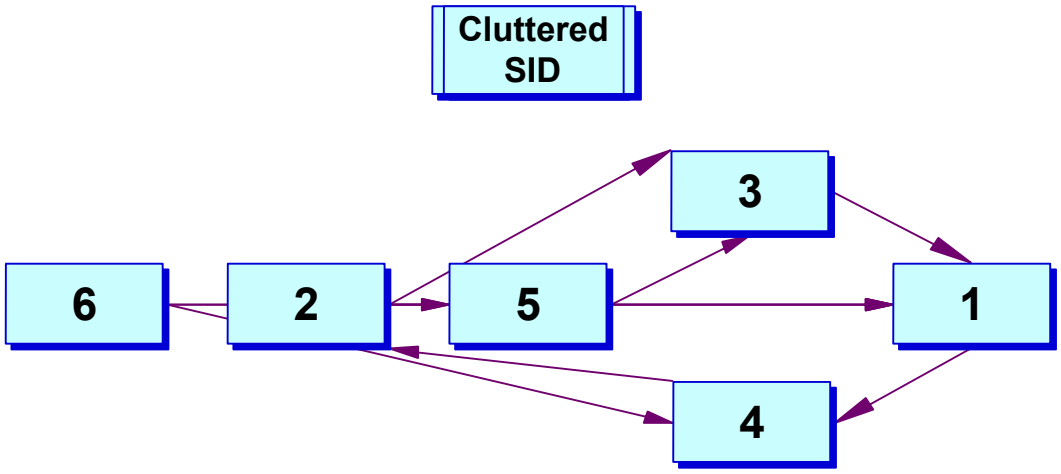
Table 3.7 Tentative SID Assignments	
6	Primary Driver
2	Secondary Driver
5	Circulator / Pivot
3	Secondary Outcome
4	Secondary Outcome
1	Primary Outcome

The delta sort represents the difference between *Outs* and *Ins*. The final version of the sorted IRD will be in descending order of delta, subject to the Zero *Outs* and Zero *Ins* rules:

- Affinities with zero Ins will always be at the top of the list, regardless of their delta value.
- Affinities with zero *Outs* will likewise be at the bottom, regardless of their delta value.

Next, the affinities are arranged in a circular pattern. Each arrow is drawn from the affinity that is the cause to the affinity that is the effect. The SID is constructed by working from left to right and top to bottom. This called the cluttered SID.

Figure 3.3
Cluttered SID

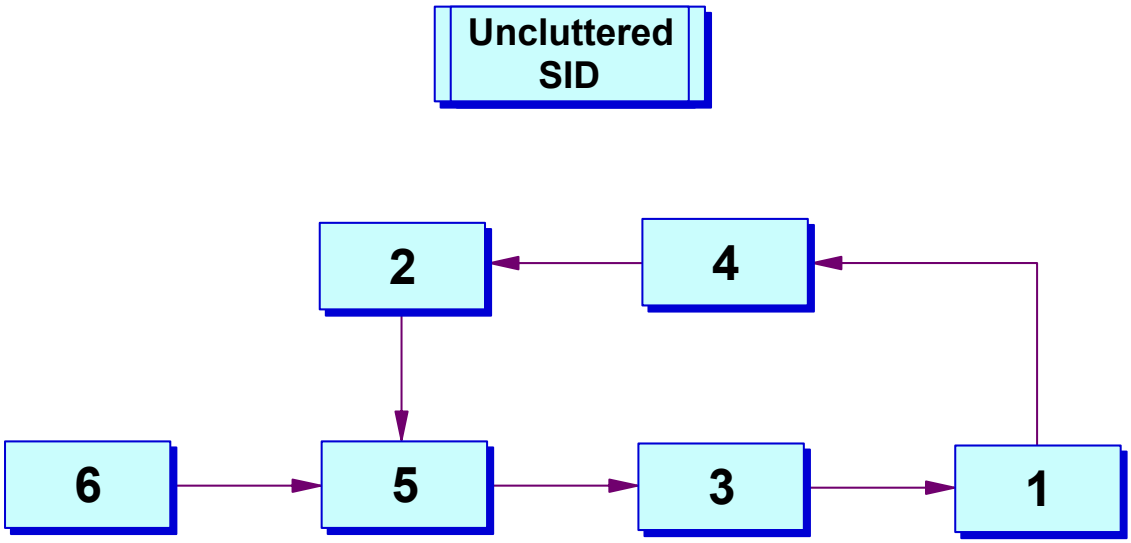


Source: Northcutt and McCoy, 2002

Next, one must eliminate redundant links and produce the uncluttered SID, a clean system. Redundant links can be thought of as the “paths of least resistance.”

According to Northcutt and McCoy (2002), redundant links are removed according to their delta and SID assignments, which is to say, the analysis begins by comparing affinities at the extreme left and the extreme right, then working back to the left. The relationship between the highest positive delta and the highest negative delta is examined. If there is any path between the two deltas other than the direct link, that link can be removed. Next, the relationship between the highest positive delta and the next highest negative delta is examined. If there is any path between the two deltas other than the direct link, that link can be removed to produce a more readable mind map.

Figure 3.4
Uncluttered SID



Source: Northcutt and McCoy, 2002

It is the simplest possible representation. These will produce results consistent with all the relationships contained in the IRD.

Step 12 – Constructing an SID from the Composite Interview Data

Incorporating all the foregoing guidelines, the study also used the Pareto Protocol to construct a SID from composite interview data**.** For each potential combination, there are three options: **A** influences **B**, **B** influences **A**, or there is no relationship. The option attracting the plurality of votes is recorded in the ART. Those relationships that attract very few or no votes are excluded from the ART. This protocol is faster, but should be used when the primary purpose of the focus group is simply to produce an affinity list rather than to conduct an extensive group analysis of the systemic relationships among the affinities.

Summary

Chapter III describes the purpose, research paradigm, study design, and methodology used to explore complex interactions, processes, and resulting student outcomes related to the implementation of large-scale assessment (high-stakes testing) policy at the secondary level. The primary focus of this study is to determine, from teachers' and students' points of view, the unintended consequences of high-stakes testing on high school students, specifically those side effects of high-stakes testing for economically disadvantaged and ethnic minority students.

Qualitative research was chosen, for it enables researchers to study social and cultural phenomena. A policy research framework was selected for the study because it allowed for the examination of multiple dimensions of a policy and to make recommendations that can help other students and schools to resolve, reduce, or eliminate the problem of unintended consequences of high-stakes testing.

Case study research in education seeks to understand specific issues and problems of practice, using the researcher as an instrument. To investigate the object of the case study in depth, researchers should use a variety of data-gathering methods to produce evidence that leads to understanding of the case and answers the research questions. According to Stake (1995), Simons (1980), and Yin (1984,) a key strength of the case study method involves using multiple sources and techniques in the data gathering process; thus, data is collected from a variety of sources.

This research is organized around three related research questions. The focus is: How do these unintended consequences of high-stakes testing affect

students (graduation, student retention)? The study's presupposition and proposition were described to promote clearer understanding. Site selection criteria are identified as criteria for subject selection.

The Interactive Qualitative Analysis method was used to undertake data collection and analysis. The IQA system allows the comparison of individual and group realities. Two major questions are answered in the research process: (1) identification of factors or affinities; and (2) identification of relationships among factors. Understanding the IQA system means identifying elements, describing the relationships, and understanding how the elements and relationships interact.

An IQA study begins with a focus group, and its first job is *silent brainstorming*. The facilitator invites focus group members to participate in a group brainstorming session. Next, the researcher facilitates a data clarification process by reading each response aloud. Third, *affinity group* or *inductive coding* occurs.

Identification relationships among factors represent all the relationships among the affinities. Output of the focus group hypothesizing activity is summarized in an IRD. An IRD is a matrix containing all the perceived relationships in the system. The IRD depicts the relationship or lack of relationship between affinities.

For deeper clarification, interviews are conducted. IQA is a system approach to qualitative research. The content of the interview is determined by the affinities developed by the focus group. The research shares the focus group definition of each affinity. The IQA interview is a semi-structured interview; the interview questions are designed and based on the affinities and sub-affinities developed by the focus group members.

The affinities produced by the focus group are used to create an interview protocol. The interview protocol is used to confirm affinities created by the focus group and to elicit descriptions of relationships among the affinities. The purpose of the protocol is to use the affinities identified through focus group data collection and analysis to inform and shape questions for the second round of data gathering: the interview (Northcutt & McCoy, 2002). The focus group data collection served as research for the interview, as well as a pilot study, providing a snapshot of the group mind-map.

The interview protocol consists of two parts: an Axial Interview—an open-ended interview design to provide rich description of affinities by the respondents—and a Theoretical interview constructed to identify relationships between affinities.

Typical interview procedures include: preparing for the interview, opening the interview, axial interview, theoretical interview, interview wrap-up, debriefing; and data analysis for IQA involves Axial Coding and Theoretical Coding. The group *Interrelationship Diagram (IRD)* summarizes the results of group theoretical coding, answering two major questions: How will the group be organized for analysis of relationships? How will a group composite (the system that represents the entire group) be constructed?

Next, a SID is made or created by taking the information found in the IRD and presenting a system representing affinities and relationships among affinities. The SID is the central visual component in the IQA System. Affinities are identified as: Primary Drivers, Secondary Drivers, Secondary Outcomes and Primary Outcome, based on their delta value.

Chapter IV provides and displays data collected from teachers and students via focus groups and interviews. Hopefully, the data will be a starting point to dispel myths about who can learn different levels of content, and to build dissatisfaction among the school community about the low educational outcomes for many students.

CHAPTER IV – DATA FINDINGS

> Organization swimming in many interpretations can then discuss, combine, and build on them. The outcome of such a process has to be more diverse and richer in the sense of what is going on and what needs to be done.
>
> (Margaret J. Wheatly, 1992, p.65)

Introduction

High-stakes testing policies assume that the testing helps raise educational standards and classroom instruction for all students and, in turn, leads to higher individual performance. This study explores the consequences of high-stakes testing (assessment) for high school students; specifically, the unintended side effects of high-stakes testing for economically disadvantaged and ethnic minority students from the critical perspective of teachers and students. Study findings are documented and presented in this chapter, with the goal of exploring the critical views of students and teachers.

System Elements

The Statement of Problem

Assessment and accountability are now prominent features of the American educational system. High-stakes assessment is a crucial part of accountability. Many students begin high school dreaming of successfully completing state exit exams and graduating. However, passing state high school exit exams is becoming the single most important hurdle for poor, African-American, and Hispanic children. Studies of high-stakes testing often overlook the unintended consequences of these assessments. However, few studies have

rigorously examined the consequences of high-stakes testing from teachers' points of view. According to Amrein and Berliner, "It is now time to debate high-stakes testing policies more thoroughly and seek to change them if they do not do what was intended and have some unintended negative consequences" (Amrein & Berliner, 2002). Thus the unintended consequences of high-stakes testing on high school students, specifically economically disadvantaged and ethnic minority students, need to be explored.

Purpose of Study

The purpose of this study is to determine the unintended consequences of high-stakes testing on high school students—specifically, the side effects of high-stakes testing for economically disadvantaged and ethnic minority students—from students' and teachers' points of view. It is important to study these unintended consequences for students during large-scale assessments for the following reasons: (1) High-stakes testing policy is acclaimed as the legislated instrument to improve teaching and learning—the ultimate goal of education reforms—by making schools more accountable for the performance of their students. (2) The possibility exists that high-stakes testing policy may be a symbolic response to very real educational problems (Ellwein et al., 1988). (3) The need exists to examine how policy made at the macro-political level (state and federal level) affects students at the micro-political policy level. (4) Legislated policy responses to educational problems often occur without sufficient study of the efficacy of the policy, its effects or the appropriateness to identify the problem (Ellwein et al., 1988). (5) Fairness and equity are needed for all students. Further, strategies to eliminate and/or reduce the unintended consequences of

high-stakes assessment to ensure all students meet and master educational standards given by teachers will be examined.

Research Questions

The research is organized around three related research questions. These questions are:

- What critical perspectives do teachers (and students) have about the effects of high-stakes testing?
- How do intended and unintended consequences of high-stakes testing affect students?
- How do intended and unintended consequences of high-stakes testing affect instruction and classroom practices?

Participants

The three focus groups included high school students who had not passed the Texas State Assessments (Group I); high school students who had passed the Texas State Assessments (Group II); and high school teachers (Group III). Students were not informed of the TAAS/TAKS status of other students in the study. All participants worked or attended an exemplary high school.

Group I

Focus Group I included eight (8) students in grades 10, 11 and 12; four African-American males; two (2) Hispanic males; one African-American female; and one Hispanic female. Seventy-five percent of the students received free or reduced lunch. Twenty-five percent of the students reside in middle or lower-middle income homes. Students in this group had failed the TAAS or the TAKS assessment in one or more subjects and needed to pass these assessments to

receive a high school diploma. Additionally, some students had failed one or more grades. At the time of the study, all students had the required academic credits to be classified as sophomores, juniors, or seniors. This group included two (2) sophomores, four (4) juniors, and two (2) seniors.

Group II

Students in Group II were students who had passed the Texas State Assessments Exams. The group included ten (10) students in grades 10, 11 and 12. Some students in this group reported passing the Texas State Assessments Exams with ease. Most students were honors students. The group consisted of ten (10) students: four (4) males and six (6) females. In terms of the ethnic background, the group was composed of four (4) African-Americans, five (5) Hispanics and one Anglo. More than half of the students received free or reduced lunch. Categorized by school class, the focus group included two (2) sophomores, six (6) juniors, and two (2) seniors.

Group III

Group III consisted of ten (10) teachers, of whom there were eight (8) females and two (2) males. Of these, eight (8) were Anglo, one was African-American, and one was Hispanic. Courses taught by the participants included Chemistry; Integrated Physics and Chemistry; Social Studies; Language Arts; various Mathematics courses; Spanish; and Biology. Teachers taught advanced, average, and below average as well as special-needs students. Teachers had an average of 15 years of teaching experience. They were also well-versed in the requirements of the State of Texas accountability system.

Data Collection

The IQA system offers the greatest possible assistance in data interpretation. Focus groups were convened to identify themes called affinities, comprising their experiences with the Texas State Assessments. Interviews and follow-up phone calls expanded on the descriptions of the affinities. Finally, pictures of systems were developed and compared. IQA reports developed for this study addressed three goals:

- Naming and describing the themes of the system
- Explaining relationships among themes of a system (system dynamics)
- Comparing systems

Each focus group began with index cards and magic markers being passed out to participants. The focus group was informed about the nature of the research, and told of the researcher's interest in their experiences with the Texas State Assessments and Accountability System. Furthermore, participants were assured that this process would identify valuable details of common themes about their experiences.

Participants were asked to select a comfortable place to sit, close their eyes, relax by taking deep breaths, and put aside duties and thoughts about today. They were asked to reflect on their experiences while preparing and/or taking the State of Texas graduation exit assessments. What memories did they have? What emotions and thoughts were involved? What were their impressions about the state assessments? They were then given the issue statement, *Tell me about your experiences preparing and/or taking the Texas State Assessment for*

graduation. The focus group warm-up used by the facilitator is shown in Appendix A.

Group Participation Process

The guided imagery process continued for about 5-6 minutes. Participants were asked to reflect on their experiences. The group was asked to think of words, phrases, mental pictures and/or memories of experiences about Texas state-required assessments, as the facilitator reassured the group that there were no right or wrong answers.

Participants were asked to take index cards and a magic marker. Next, individuals silently "brainstormed," writing their thoughts and reflections on index cards, using single words, short phrases, or even diagrams that came to mind regarding the issue statement. Focus group members were encouraged to produce as many cards as they wished. Participants were assured that whatever they wrote on the cards would stay confidential, and that all cards would be put together so the author would not be known. When it appeared that the group had generated as many ideas as possible, the cards were collected and taped on the walls in no particular order.

After the cards were taped on the wall, the focus group was asked to silently read the cards. After the group was finished, the facilitator went to the wall and read each card aloud, allowing the group to arrive at a socially constructed, shared meaning of each card as decided among members of the group. The group was asked if each card made sense. If the card was not clear, the author or a member of the group was asked to explain what it meant, clarifying the meaning of the card.

Next the group was asked to silently move the cards into columns of similar themes. If they disagreed with the placement of a card, they were free to move it to another column. This process continued for several minutes until members of the group agreed with the placement of each card. However, some individuals within the group seemed to get frustrated when a card they had placed was moved. They were requested not to talk about where the card should go and were assured that if the card's meaning and placement were ambiguous, the ambiguity would be cleared up in the next step.

After the process of meaning clarification and card placement, the group was given the opportunity to write out any new cards that might have come to mind after reading what other members in the group wrote. Several new cards were generated and the authors were asked to place the new cards in the appropriate columns.

The columns were given names. The facilitator began with the column that seemed to be the easiest to name, and asked the group to give it a name that would reflect all the cards in the column. A new card representing the name was placed above the column. The process was continued until each column was named. The named columns were reviewed to see if any columns might be combined. Some columns were combined under one newly named category; the original columns became sub-affinities of the new affinity. The cards were reexamined to ascertain if they were still in the correct category or now belonged in a different column. If a category was too complex, the group was asked to give it a sub-category if necessary. Sub-categories were also given names. After the cards had been arranged and the affinities named, the affinity production exercise was complete.

Affinity Identification and Description

Group I – Affinities and Descriptions

In response to the issue statement, Group I (students who have failed TAAS/TAKS) generated more than 100 responses in the form of a word, phrase, or sentence on 5x8 cards. As previously described, the cards were then sorted according to common themes called affinities and were given identifying names by the focus group. The following are seven themes that emerged from the focus group of students who had failed the Texas State Assessment together with descriptions as expressed by the group discussion and interaction. Affinities chosen by Group I included:

- All or Nothing
- Emotions
- Dropouts
- More Time to Learn
- Practice Tests
- Tested Materials
- Use of Class Time

Descriptions of these affinities are as follows:

All or Nothing. The group described *All or Nothing* as the State of Texas's graduation policy—i.e., to receive a high school diploma, students must pass the state's assessment exams.

Emotions. *Emotions* represented how the group felt about preparing and taking the State of Texas Assessments for high school graduation.

Dropouts. *Dropouts* identified students who have left school before graduation due to academic or state assessment failure or students who

completed required courses for high school graduation but failed the state assessment.

More Time to Learn. This affinity reflected students' desire for additional time to master State of Texas core requirements.

Practice Tests. This affinity included district-wide TAAS/TAKS practice tests; pre-tests for the district TAAS/TAKS practice test; benchmark testing; and regular tests for courses.

Materials Tested. Students identified materials tested as specific concepts examined on the test, and when or whether these concepts were taught during their middle and high school career.

Use of Class Time. This affinity refers to the instructional focus of day-to-day classroom instruction and activities.

Group II – Affinities and Descriptions

Group II comprised students who have passed the TAKS or TAAS exams. Students were given the issue statements, and then generated more than 90 responses in the form of words, phases, or sentences on 5x8 cards during a silent brainstorming session. The affinities identified by this group included:

- Motivation
- Too Much Weight on the Test
- Emotions
- Benefits
- Equity
- What I Want to Learn
- Dropouts

Descriptions of the affinities are as follows:

Too Much Weight on the Test. The group identified *Too Much Weight on the Test* as a State of Texas graduation policy—i.e., to receive a high school diploma, students must pass the state's assessment exams. Failure to pass the assessment meant academic failure to successfully complete high school.

Motivation. The students described this affinity as a stimulus or desire to learn course materials and prepare for state assessments.

Emotions. *Emotions* represented how the group felt about preparing and taking the State of Texas Assessments for high school graduation.

Dropouts. *Dropouts* identified students who have left school before graduation due to academic or state assessment failure or students who complete required courses for high school graduation but fail the state assessment.

Benefits. This affinity reflected positive perceived outcomes of the Texas State Assessments.

Equity. This affinity included district and state policies that address the educational needs of students. Students who had passed the state assessment perceived that schools and teachers were not doing enough to help them develop academically.

Use of Class Time. This affinity refers to the instructional focus of day-to-day classroom instruction and activities.

Group III – Affinities and Descriptions

Group III comprised teachers in the study who were members of an exemplary high school staff. As previously noted, courses taught included Chemistry, Integrated Physics and Chemistry, Social Studies, Language Arts, Math, Spanish, and Biology. Teachers taught advanced, average, and below-

average as well as special-needs students. Teachers had an average of 10 years of teaching experience; in addition, they were well versed in the State of Texas Accountability System.

The affinities or themes reported by this group were:

- TAAS/TAKS Assessment Graduation Value
- Instructional Focus
- Practice Tests
- Emotions
- Motivation
- Dropouts
- Bias
- Teacher accountability
- Class size/Resources

Descriptions of affinities identified by this group included:

TAAS/TAKS Assessment Graduation Value. The group described this affinity as a State of Texas graduation policy, requiring that, to receive a high school diploma, students must pass the state's assessment exams.

Instructional Focus. The affinity *Instructional Focus* describes instructional practices and curriculum decisions based on the school's need to receive a successful rating from the Texas Accountability system, and how these changes affect student learning.

Emotions. *Emotions* represented how the group felt about preparing students to take the State of Texas Assessments for high school graduation.

Practice Tests. This affinity included district-wide TAAS/TAKS practice tests; pre-tests for the district TAAS/TAKS practice test; benchmark testing; and regular tests for courses.

Motivation. The teachers described this affinity as the stimulus or desire students processed or the teachers created within students to learn course materials and prepare for state assessments.

Dropouts. *Dropouts* identified students who have left school before graduation due to academic or state assessment failure or students who complete required courses for high school graduation but fail the state assessment.

Bias. This affinity reflects cultural and language barriers inherent in instruction, class materials, and Texas graduation assessment.

Teacher Accountability. This affinity represents State of Texas inferred standards of excellence for all teachers, requiring teachers to teach state standards and core objectives for all students, including students considered "low achievers."

Class Size/Resources. This affinity represents resources provided for the educational development of students, including class size.

Interview Protocol

The interview protocols were derived from the reconciled affinity list. Brief description is used to describe the affinity so as not to influence the responses given during the interview. The interview protocol consists of two parts: 1) the open-end *axial interview* designed to provide rich description of affinities by the respondents; and 2) the structured *theoretical interview* designed to identify

relationships among affinities. The axial interview is addressed in this section. The theoretical interview will be addressed later.

The axial interview described the common themes or affinities identified by each of the three groups. Axial interview protocols for each group are shown in Appendix B - D.

Transcripts and Axial Code Tables

Interviews were transcribed and analyzed for *axial codes*—i.e., specific examples of discourse that illustrate or allude to an affinity. The reference was documented for retrieval by recording the affinity number on the line of the transcript that refers to the affinity and by documenting the line numbers and affinity numbers in the individual interview Axial Code Table (ACT). Next, quotes relating to a specific affinity were retrieved and inserted into the third column of the ACT, along with the line(s) of the transcript that were the source of the axial quote, which were also posted. After all interviews had been coded, the data from the interviews were summarized to create a composite of the individuals' experience with the phenomenon. Axial data were transferred from each individual interview Axial Code Table to a Combined Interview Axial Code Table. Combining all interviews into one table for each group allowed the researcher to create a database for the entire set of respondents containing all axial codes for all affinities, with each code containing a link or a reference to the transcript and line numbers that produced the code.

Composite Affinity Descriptions

Quotes were examined for each separate affinity. The quotes for a particular affinity were organized into sub-groups. These subgroups contained

quotes that addressed a common theme describing that affinity. Multiple quotes were then woven together to develop a composite quote (which is included in quotations in the following discussion of each item). The following section is a composite description of the affinities based on quotes obtained from all the interviews.

Group I – Composite Affinity Descriptions

Quotations from Group I (students who have failed the TAAS/TAKS) were examined in order to determine how the students perceived the Texas State Assessment Policies affected their learning and performance on the state assessment. The students expressed their understanding of the assessment graduation policy, as well as how this policy has the potential of changing their educational futures and careers.

All or Nothing

I don't think the test should determine everything. For this group of students, the reliance on just a simple measure to determine who will receive or who will be denied a high school diploma was not valid because it did not adequately test their knowledge. "At the end *what it comes down to is if you pass the test or not*. If you don't pass the exam, then all of the work you have done has been wasted. I know a lot more than what was on that test." Most of the students think that schooling is more than merely the results from one test.

The test does not measure all I know. **"***Forty multiple questions on a test do not measure what I know*. I know a lot more than what they asked on the state assessment. I wish the test had another format like short answer, or they would use a portfolio to see what you can do. I know a lot, but I fail the TAAS."

School should be more than a test. The students questioned the value of school within the context of state assessment policies, and the multiple-choice testing instrument used.

> Do you not think school should be more than just a test, or if kids can pass a certain test or not? Is the student who passed the test by one question more intelligent than the kid who failed the test by one question? I do not think so…it is just luck sometimes. And more than one or two questions should determine who graduates. Kids sometime do better on different types of tests. I am the kind of kid that does not do well on multiple-choice tests. I would just like them to ask me a question straight out and let me answer it. In class, when we are answering questions, I do okay.

The penalty for failing the test is too high*.* The students addressed the consequences of failing the state assessment and the potential impact to their future.

> *You work hard for 12 long years, and for some kids it is really 13 or 14 years* because they have been held back. You know they failed a grade or two. And if you fail the test you do not graduate. All the work was for nothing. It is not fair. What kind of future do you have if you do not have a high school diploma? What kind of job can you hope to have? I can tell you . . . you can hardly get a job at a fast food place for very little money. How can you get anywhere on that kind of a job? *I do not think one test should destroy someone's future.* What do they think we are going to do? *Without a high school diploma you cannot make an honest living*. You need it for the basic things in life like food, clothes, and a place to stay. You cannot get a job without it.

The test is not exactly fair*.* Here, students are expressing their opinions that state assessments are not fair and do not yield valid and reliable score interpretations, noting language bias.

> This testing stuff is not exactly fair. I do not think one test should decide what you know. There is no way you can make me think that the test is written is such a way it can tell who should or who shouldn't graduate from high school. Sometimes the questions on the test are not written in the English we use in class—what I understand.

Emotions

Emotions referred to feelings students experienced preparing and taking the state assessment. They expressed how they felt about the experience. Most of their experiences were negative.

So Much Pressure! Students are expressing the depth of their feeling for the experience of high-stakes testing, using the term *pressure*.

> There is so much pressure about passing the test the first time. But, if you have failed it before, *the stress and pressures to pass the exam are almost unbearable*. I understand why some kids just give up and drop out of school. I have often thought about dropping out. Teachers put too much pressure on us about passing the test. *Some teachers are always saying if you do not do well on the test you will not graduate*. It is not like I do not know this already. I am really trying. There is just a lot of pressure about the exam.

I cannot make you understand how much stress I have about taking the TAKS. Once again, students are expressing the timbre of their experiences with high-stakes testing***.***

> If I fail it again what do I do? I cannot make you understand how much stress I have! There are months before the exam and I am so nervous and full of fear about taking it. It gets very stressful when it gets close to the test time. I am so nervous I can hardly sit still for the test. I fear that I will fail it. The test freaks me out. It is all or nothing. I do not know how I am going to face my family if I can't pass the test and get a high school diploma. Just thinking about how the test affects my future, freaks me out.

I fear that I will let my family down again*.* This is greatest negative outcome expressed by the students.

> What will I do if I can't graduate? They [family] are all looking forward to my graduation. My dad was so hurt when he found out that I had failed the TAAS. And when I failed the TAAS exams, I felt that I had let my family down. There was little hope for college or my dreams, and I had very little hope for college or a good job.

Dropouts

Dropouts identifies students who have left school before graduation due to academic or state assessment failure, or students who complete required courses for high school graduation but fail the state assessment.

You work so hard. These comments express efforts exerted by students while preparing for the state assessment.

> Tutorials, worksheets, assignments, practice tests, and anything else you think of will help. *Some people have already failed a grade but they keep trying.* However, after a while it gets to some kids, and they just stop coming to school. How do you face your friends and family after you fail the test? Some kids see no way out.

In the following comments, students are expressing their need for additional instructional support and remediation.

> They should *do more* to help people who are really thinking about dropping out of school because of the TAKS exams. The school, the state*, somebody should do more to help students* who were struggling to pass state assessments, especially those students who came to school on a regular basis and do the class work. Last year 1% of our student body failed the graduation test. This year they are saying it is going to be 30% of the student body. *When I hear things like this I feel hopeless.* I think about giving up.

Students dread the possibility of being labeled. “I failed the test and I know my folks are disappointed.

> *I do not want to be labeled stupid.* I try to tell myself I will do better next time, but I am not sure. I know what people will say. I try to play it off, but it hurts. You do not ever get over this kind of thing. Your family and friends look at you differently. No one wants to be thought of as being stupid.

More Time to Learn

More Time to Learn refers to additional time the students desired to learn materials or concepts required by the state.

Middle school should do a better job in preparing us for some high school courses. Students perceived that their school system had not given them the opportunity to learn core materials required for high school. “Some of the things my teachers in high school think I should have learned in middle school I have never heard of. I was not taught those things. Maybe we should have done more in middle school? When were we supposed to learn the stuff? Hey, I did not think middle school was that big a deal.”

Take time and teach us stuff we will need later in life. Students in Group I questioned the use of instructional time, the pace of their class, and concepts they were taught.

> Don’t teach us stuff or have us memorize stuff just for a test. I do not like learning things I will never use after high school. Some teachers start too late teaching us, and they don’t teach us much**.** Other teachers just try to go over and over what they think is on the TAAS/TAKS test. I think they should start earlier in teaching the things we need to know, and not with just a lot of worksheets. We need time to learn the important things. Some teachers wait until the last minute. Then they rush the class, saying how much more we need to learn before the state test. And it causes students to work harder and try to learn faster, but that is not right. Some kids can’t keep up. It gets tough for students who are not as fast learners. I need more than hit-and-miss tutorials. Some people just need more time to learn. They should start a program for kids who do not graduate with the TAAS test the first time. I would not mind. I need more time with certain things. I am not sure how to come back to high school after you have successfully completed all of the course requirements but you have not passed the TAKS, and to keep trying for the diploma.

From the students’ viewpoint some teachers were not willing or equipped to teach low-achieving students.

> They should really teach kids. Some teachers feel it is a waste of time to teach kids, like me, who may not pass the TAAS/TAKS**.** There should be more help for students who are not Special Ed but are not “fast” learners. All teachers write, read, grade, but sometimes some teachers do not know how to, or cannot, help kids who are “slow.” They don’t get it. Some of them really try. We need time to really understand. But other teachers take

> time and explain things in ways everyone can learn. *They are special.* It helps when you get a good teacher.

According to the responses, high schools should become more like colleges, where a student could choose programs and/or how many semesters he/she needs to complete the work. The students gave their solution to rectify students' failures of state assessments. "*I needed more time in courses like science and math*. Instead of having to finish some course in one year or semester, I would like to have more time and go slower, instead of failing. Maybe a longer class time for some classes during the day."

Practice Tests

Practice tests included district-wide TAAS/TAKS practice tests; pre-tests for the district TAAS/TAKS practice test; benchmark testing; and regular tests for courses.

Too many practice tests. In the opinion of the students, there was an excess number of practice tests.

> We have far too many practice tests**.** Many times we were only studying sample items for the TAAS/TAKS. *Sometimes I am not quite sure how some questions related to the course I am taking*. I just try to learn what I can and hope I pass the TAAS/TAKS. *I am numb for taking so many practice tests.* School-wide practice tests, objective practice tests, sample practice tests, teacher-made-up practice tests, practice tests in tutorials, on-line practice tests, practice tests over and over again. [Laugh] I am becoming a good test-taker. *There are so many practice tests, you can't take them all seriously.* Often it appears that the only thing that is important in school is passing the test, by memorizing unrelated facts.

Material Tested

Students identified *Material Tested* as specific concepts examined on the test and when or if these concepts were taught during their middle and high school career**.**

I do not think I was ever taught half of the things I saw on that TAAS test. Students questioned whether they were given the opportunity to learn the information assessed on the exam.

> When I asked one of the teachers about it, she said I was to have learned it in middle school. She knows she did not cover that stuff in class. I do not want anyone else to say I should have learned this in middle school or elementary school again. That is not the point. I am here. I did not get it. I need to pass the TAAS. And anyway, the teacher should teach me.

The students were concerned with the complexity of information on the state assessment as well as when it was taught.

> *Some test items were very specific over concepts we covered years ago and had very little to do with our current courses.* For example, the TAKS for Science had specific items on it about the human body that were only covered in middle school. They were not covered in Biology, nor were they reviewed during our review for the state exam. Some of the things on the test I remember being taught in middle school, but not in high school. It was so hard trying to remember them on the TAAS test.

Use of class time

This affinity refers to the instructional focus of day-to-day classroom instruction and activities.

I want to learn things I like and I feel I need that are not on the TAAS/TAKS test. The students expressed a desire to learn additional concepts they consider essential.

> Sometimes we just go over stuff the teacher says will be on the test. It is boring. Sometimes we are talking about things that really could help you understand; things not on the TAKS but [which] are important for real life. The *teachers know information but we do not have time to discuss it in class because it is not on the TAKS*. Sometimes I wonder if I could just make up a guide of things I want to learn and go study them. Or, if our teachers are too busy with the TAKS, they would let some kids take a virtual course or some kind of computer-interactive course.

Life involves more than a few questions for TAKS. The students wanted to determine part of the curriculum they studied. “Everyone is not going to college. Kids should have a say in what they want to learn. I want to do more in class than a lot of worksheets or drill-type things. I want to find something I am really good at.”

Group II – Composite Affinity Descriptions

Quotations from Group II (students who have passed Texas assessments for graduation) were examined to develop composite affinity descriptions.

Motivation

The students described this affinity as a stimulus or desire to learn course materials and prepare for state assessments.

Sometimes I do not feel motivated. Students who have passed the state assessment or students who are sure they will pass the state assessment are not interested in topics covered in class.

> I wish my school and teachers would feel more confident and less stressful about the TAAS/TAKS. It is easy to be motivated when you can’t see a reward for your work and efforts. However, sometimes in school we are not learning new things. And I do not see the point of the test. And once you know you can pass the state test or you have passed the test for graduation, you are no longer afraid of not graduating. Nor are you motivated to study hard for your courses.

Too Much Weight on the Test

This affinity expressed the students’ view that the Texas state assessment should not be the only measure to determine if a student was granted a high school diploma.

I think that is necessary for all kids not to be promoted until they have passed the state assessments. Here students expressed the need for

state assessment. "*Many kids are not qualified to graduate and I do not think they should.* If they pass kids without their knowing basic things, the next year is harder for them. I really do not like the state tests, but now all jobs require a certain level of knowledge so we might as well get started somewhere."

Emotions

Emotions represented how the group felt about preparing and taking the State of Texas Assessments for high school graduation. Many of the emotions associated with the state assessment are negative. Additionally, the students indicated they experience pressure and were concerned about their test performance.

> I understand why we take the TAAS/TAKS, but the pressure applied is not necessary. Some of the smartest people break down when it comes to taking tests. I don't think you should be totally denied a diploma if you don't pass. It is not fair to many people because it gets them really worried about what is on the test—when the school is not sure what is on the test. Most of the questions on the test we do not know. So you guess.

Usefulness/Benefits

This affinity reflects perceived positive and negative outcomes of the Texas state assessments.

I think the TAKS test was very useful in certain instances. Here, students identified positive benefits of the TAKS assessment.

> It makes certain that teachers are teaching all students some basic skills. It also sets standards for students, teachers, and the school to live up to. The test encourages standards. And it prepares us better. We really know how to pass those types of tests. But I don't feel there is enough time for my pre-AP and AP classes. I really do not like the state tests, but now all jobs require a certain level of knowledge, so we might as well start somewhere."

Other students viewed the state assessment as being neither very useful nor reasonable.

> The test is pointless because TAAS seems to be *repetitious.* All the questions seem to be the same every year. Even the sample questions are the same or similar. Some teachers just focus on the TAAS/TAKS test questions. Simply put, *I feel that it is pointless—we go to school and learn how to take a test. And high-stakes testing* is *not really that reasonable*. Because most of the stuff you learn the whole year is all based on something you already know. So why are we taking all these classes, when in the end it comes down to the test? It is simply not fair. Seniors shouldn't be penalized and kept at school just to experiment with the state test, especially if you have already passed the TAAS assessment for graduation. We should not have had to take the field test. *It is not fair! It does not affect us!"*

Equity

This affinity included district and state policies that address the educational needs of students.

Sometimes we are tested on concepts that we have not covered in class, or concepts that we learned in earlier grades. Here, students are concerned about the length of time between when a concept is taught and when the concept is assessed on the state exam.

> Too much time has elapsed between when material or concepts were taught and when we are tested on specific concepts. I do not think the whole year's work should be based on something we have not been taught that year. Seniors should not have to take a practice TAKS test, nor the TAKS, during this year of transition, since TAKS did not affect whether or not we will receive a diploma. We have to pass the TAAS.

I realize that a lot of time, effort is needed to be sure that all kids know basic stuff. Students indicated they are concerned that they are not being taught advanced topics and/or concepts.

"But I think the school should do more for kids who have mastered these things and have teachers should teach more than the TAAS/TAKS test. I think equity goes both ways."

What I WANT to Learn

This affinity refers to the instructional focus of day-to-day classroom instruction and activities. The state assessment was seen as the main focus of instruction in the classroom, and preparing for it was viewed by students as being distracting and prohibiting and/or interrupting "real" learning.

> I think throughout the year the main focus in most classrooms is the TAKS test. *A lot of teachers are reluctant to teach concepts or discuss ideas they think may not be on the TAKS exam.* Practicing for the TAAS/TAKS test is sometime distracting. When we are learning things I want to learn we have to stop and practice for the test. I wish we studied more things.

Dropouts

The *Dropouts* affinity describes the tendency for some students to leave school when they fail the TAKS or TAAS exams. The Students stated reasons they perceived students drop out of school or why they have considered dropping out of school.

> I knew a kid who was so disappointed by failing the TAAS test. He did not want to come back to school. I know if I had failed the TAAS/TAKS I would be so disappointed. I have a friend who has not passed the test. It looks a bit hopeless for him. If he does not pass it, I think he will not come back to school. It is very embarrassing when you fail the TAKS or TAAS tests, especially if you are in honors classes and you realize that, if you do not pass the state assessment, you will not graduate from high school and go to college. Some students are just depressed and refuse to come back to school. Others are so nervous and frustrated about failing the test they cannot seem to settle down to take the test again.

Group III – Composite Affinity Descriptions

Quotations from Group III (high school teachers) were assessed and selected to derive composite affinity descriptions.

TAKS Assessment is Weighted Too High

This affinity identifies the beliefs of the teachers the state graduation assessments test should not be the sole criterion used to determine if students received a high school diploma. No one assessment should determine if a student graduates from high school with a diploma.

> The consequence for failing the TAKS exam is too high for students. We spend 12 or more years preparing students, making sure they know the state standards and objectives. Then the state says, "Oh by the way, if you do not pass the test for graduation, you will not get a diploma." I cannot say that this is a fair consequence for these students. It has been my experience that many of my students who failed the Texas state assessment fail by one or two questions. And when "at-risk" students I have taught and worked with fail the Texas state assessment by one or two questions and do not receive a diploma I feel like a failure. And to say that a student should not get a diploma based on this one assessment is not reasonable and it is unfair to students and teachers. *The State and policymakers see these students as numbers. I see them as Adam or Melody.* Policymakers take all of school and say the only thing that really counts is the test. It is not the best way to do things. I am not bitter. I do not want you to think that. But I know that this is not the best thing for students. *It does not promote higher learning.* Promotions and graduation should not be linked to only one measure.

In some ways I agree that some people have had a hard time with the TAAS/TAKS test. Here teachers discuss the challenges certain students have with multiple-choice exams. "The TAKS test has a multiple-choice format, and some students do not test well using this format. They are different kinds of learners. Some kids just can't take test. Those kinds of kids are failing through the cracks. We should acknowledge the damage high-stakes testing does to some children."

A student can fulfill all of the state requirements for coursework – yet not get a diploma. "Is that in the best interest of students? *I think as educators we might go for years and years without realizing or acknowledging what damage high-stakes testing does to children."*

I am hoping for a better system. The teachers acknowledged a need for assessment reform.

> Frankly, I would like to see a dual system with the test and class grades. I think the TAAS test was weighted too high. Students can finish all the coursework and not get a diploma. The system needs improving. It is not fair for those students that cannot pass the state assessment and have passed all course requirements. Not getting a diploma can affect a student's entire future. If a student can fulfill all of the state requirements for coursework, they should get a diploma. It is in the best interest of students and the system. In our current system, most of the consequences for success or failure remains with the student.

The Assessment is needed*.* The teachers agreed that state assessment is required.

> I agree that some people have had a hard time with the TAAS/TAKS test. They are different kinds of learners. But do we go back to social promotions? Only educating a few? I am a coach and I grew up playing sports during high school in Texas. It was not uncommon to just pass athletes through school without making them do school work. Many students did not get an opportunity for an education. The school system simply did not require them to study. As long as the school was winning games, the "good" athletes were never worried about failing classes or not graduating from high school. This unwritten policy was extremely harmful to many students who left high school without basic skills. So I see the state assessment as a necessary requirement. However, I do not think it should be the sole measure that determines if a student receives a high school diploma.

Many students did not get an opportunity for an education. The school system simply did not require them to study. As long as the school was winning games, the "good" athletes were never worried about failing classes or not graduating from high school. This unwritten policy was extremely harmful to

many students who left high school without basic skills. So I see the state assessment as a necessary requirement. However, I do not think it should be the sole measure that determines if a student receives a high school diploma.

Practice Tests

This affinity identifies the beliefs of the teachers that far too many days are required or devoted to TAKS and TAKS preparation, and describes pros and cons of practice tests from their perspective.

Practice makes perfect. Some teachers considered TAAS/TAKS practice tests to be an essential part of their instructional strategy.

> Practice tests are very, very, important! I mean full-fledged practice tests for TAAS are essential for all students to pass the Texas state Assessment, not just for "at risk" students. Students need to become familiar with the wording and type of questions on the test. In my eight years of teaching, I have found that *the more students practice for the test, the better our scores are*. Therefore, I do not have a problem with practice tests. Practice makes perfect. *The more students practice, the easier it will be when they take the real thing*. I always tell my kids *it is a necessary evil*. Practice tests are ways teachers can even out the negative aspects of the system. I see practice tests as a form of damage control until the system is improved. I, too, value practice tests. They help students get ready for the real thing. Some kids would not pass the state assessment if we did not give practice tests. I think practice tests benefit students. They help them to be aware of what to expect. The more we prepare for the test, the better the students will perform on the state assessment. But we spend more time on practice tests and preparing for them than I would like.

Too many days of TEST, TEST, TEST! Certain teachers presented challenges they have experienced with practice tests, as well as their view of the usefulness of practice tests.

> We spend too many days on testing. For this year, at the freshman level, we have had four (4) days of district TAKS tests and seven (7) days dedicated for pre-TAKS testing. *We spent too many days getting ready for the state assessment—practice tests, reviewing for the practice test, plus days lost in the educational year because students are burned out on*

testing days. The students are tested too much. Kids have had enough. They are bored and tired of testing by 10th grade. Additionally, students who had passed the TAAS are often bored with TAKS practice tests or practice TAKS preparation. To put it kindly, I despise practice tests! Despise them! My main complaint is they take up too much class time. However, in order for any school district to be exemplary it must administer a number of practice tests, often reducing valuable class time. We use 11 to 12 days for TAKS and practice TAKS tests. *We lose too much class time.* Think about it. We have about 185 days each year to teach our subjects, approximately 185 hours each year. If we spend 12 hours just on the TAKS tests and practice tests. add one class period to review for the TAKS test per week for 30 to 35 weeks, what's that—35 plus 12 [pause] . . . 47 hours? That means *a little more than a quarter of my class time is devoted to preparing or taking the TAKS exam.* It is too much. The quantities of practice tests we administer take too much time out of my teaching time. There is not enough time left to teach for mastery of all the objectives. We take up too much valuable class time for practice tests, and it has no value for students sometimes. I once suggested that we test students in the ninth grade to see where they were academically—then decide what kind of TAKS prep was needed for each student, exempting those students who did not need help. Well, the district did add the test. But everyone still gets the full practice test treatment, no exemptions. Instead of helping the problem, my suggestion added more days to the testing schedule. So, since then, I have been quiet. But practicing for the TAKS is not preparing students for post-high school. We do not teach them the higher-level skills they need. We are teaching to the test. I think some students are going to have a hard time in college.

Instructional Focus

This affinity encompasses the impact of TAKS assessments on instruction and classroom practices. Teachers described negative and positive attributes of having a statewide assessment.

Positive Attributes

Expectations. *"*Texas State Assessment allows for the establishment of common expectations for teachers and students. Expectations that are the same for all students are long overdue. Now we have general *expectations for all students***.** Teachers cannot get away with not teaching certain types of kids."

Core Curriculum. "It has caused Texas to develop a base curriculum, with stated course objectives. A common core curriculum is very beneficial. We now have statewide standards and goals for both teachers and students. That is a good thing. It also forces the majority of students to be taught a higher level."

TAAS/TAKS Prep. "When we had block scheduling we saw students every other day for about one hour and a half, and devoted 20 minutes daily to TAAS and taught our curriculum the remainder of the class. *Now one day of the week is set aside for TAAS.* Our scores have improved. A few years ago when I came to the math department, our TAAS scores were low, in the mid-fifties; now they are in the nineties."

***No longer an Island Unto Oneself*.**

> Before the TAAS, there were teachers who taught what they liked and they spent most of the year doing it. *Each teacher was an island unto him- or herself. When a student went from grade to grade, you could not be sure what was taught*—making it unfair to students at times. But with the state assessment, teachers know to teach a more common curriculum to all children, with common goals and expectations. Now, in math, we give the same quizzes, tests, reviews, and assignments. We all have our own individual teaching style, but common core curriculum is proving to be a benefit. Now, when students move from class to class or district to district, we can be certain that students have a common core of knowledge.

***Public Awareness*.** Teachers are aware of the impact of public opinion and wish to be recognized for their outstanding efforts with students.

> Like any profession, we have good and bad members, and nothing puts a move on bad teachers like having their TAAS/TAKS scores announced to the public. But TAAS/TAKS scores can help the community understand what an outstanding job most teachers perform on a daily basis. Also, we have stopped social promotions.

Negative Attributes

***Quality of Instruction*.** The teachers expressed concern that the current emphasis on the state assessment test adversely affected the quality of instruction.

> Testing leads to a scatter-gun approach to instruction . . . Hit 'em with a lot of little things; that does not encourage depth of study. I hate to say it, but all teachers sometimes teach to the test. We are teaching to the test and sacrificing objectives in upper-level courses. The TAKS test becomes a priority to teachers and students instead of the overall educational experience. The bottom line is: the TAAS/TAKS assessment and school accountability lead many teachers to teach to the test. They simply want their students to survive the system.

***Teaching to the Test*.** A recurrent complaint of the teachers was that the present system forces them to focus their teaching excessively on the state assessment test.

> I have to teach to middle and below students. And in all my classes, it is about what is on the test. I know that it is not politically correct to say it, but there it is. Students must pass the math section of the graduation assessment. I know focusing most things in classroom instruction around a test is a mistake. But no school wants to be seen as low-performing. And the system of testing and retesting has worked for our district. We are exemplary, and anyone talking about the amount of practice tests or changes made in the curriculum is not going to win friends and influence people. Therefore, many teachers are choosing the easy way out and teaching to the test. It makes my job harder. And it is harder for other teachers who try to do more. We need to stop "dumbing down" the Texas curriculum! Stop teaching to lower level! With inclusion it is easier to go down. And we teach down because it is the easiest thing to do. The bottom line is the TAAS/TAKS assessment and school accountability leads many teachers to teach to the test. They simply want their students to survive the system.

***Narrowing Curriculum*.**

Teachers are concerned about curriculum content and curriculum control within the classroom.

> The TAAS test causes narrowing of the curriculum; and teaching to the 'test.' We are cutting up our curriculum. The curriculum is controlled by the administration and not by the teachers who know best. The state assessments affect the curriculum and classroom teaching. Teachers used to have more control about what happens in their classrooms—what was taught, when it was taught, and how it was taught. Now it is the state that says what is important. Put aside what I as the teacher may think is important or will benefit the students I see every day. Most of the instruction in the classroom is centered on the TAAS/TAKS tests. Teachers are very aware that their students must do well on the test.

***Curriculum Objectives*.** The teachers are requesting that state curriculum goals and expectations be written in a clear, relevant manner, allowing easy conversion to daily lesson plans.

> The state assessment should foster equal expectations. But with the new assessment TAKS, the state objectives for the TEKS in math are not very clear. They are vague. I think before we can have clear expectations for students, we must have clearer expectation objectives for teachers to teach. "At risk" students and "bubble kids" are students who will suffer more from this policy. They could fail because teachers are not clear about state objectives. "Bubble kids" are students who are likely to fail the state assessment by one or two questions. The state needs more teachers' input about what is to be taught. Sometimes everyone is not on the same page. I try teaching all the state objectives. *But there is not enough time to teach all the objectives.* I can but there is not enough time. However, at this moment in time, teachers are now concerned about poor kids and minorities. So, on paper, we look wonderful. But sometimes, we are just training students to take the test. We have lost so much instructional time. I do not think the tests increase learning. We are just teaching students to take tests effectively.

***Meeting the Needs of Students*.** The teachers in the focus group communicated a need for one-on-one instruction time with students to effectively address their academic needs and development.

> I often do not have the time to spend that is required for low achievers or students with different learning styles to fully understand the concepts, especially in large classes. Nor do I have time to teach concepts not on the state assessment to these students. Additionally, advanced students are not given the extra time needed to help them develop beyond the state standard. We need to find better ways to teach each student in the areas

> they need most. Assessment should be used to find out what a child's needs are and to help fulfill these needs in the classroom.

Not enough time. The teachers are expressing the reality of the classroom—that students need time and attention to learn.

> I often do not have the time required for low-achievers or students with different learning styles to fully understand concepts, especially in large classes. Nor are concepts not test-taught to these students. Additionally, advanced students are not given extra time needed to help them develop beyond the state standards.

Looking for Solutions. The teachers indicated they have explored and are exploring the most efficient processes to enable students to effectively pass the state assessment.

> I took every release test apart, looking for patterns. And I found several patterns that I have passed on to my students such as "answer not found." Normally on the TAAS test there are 7-9 questions on the test with this answer choice. However, this answer is correct for three to four questions. So I taught my students to never choose this answer choice unless they are absolutely certain, and never choose more than 4 questions. I also found that there were certain TEKS objectives that our students were weak in year after year—objective 1, and objective 4. I altered my lesson plan to make sure that these objectives were taught at the beginning of the year and reinforced throughout the school year. But the most important thing I have done is to use flash cards to help students master objectives. Flash cards contain the most important things from a unit. Kids practice with each other the first few minutes of each class to see if they have mastered a concept. They also show kids that they are learning.

Preparing for the real world. The focus group teachers expressed concern that the current approach to state assessment does not adequately prepare students for realistic real-world challenges or cultivate the intrinsic capabilities of some students.

> I do not think the TAAS/TAKS focus in schools prepares students for the real world. Students learn to master the test, but what does that say about education? When I was in high school I was not the smartest kid. It was not until I went to college that I realized that I was smart, very capable, and tried to apply myself. If we had this system when I was in high school,

> I could have been one of those students who did not pass the state assessment. Schools should be about promoting learning. We must be willing to support students in and out of school. The TAKS test does not make good students. It does not make good citizens. The consequences are too high for some students. And for students, the immaturity factor is huge. Some of our students do not realize that their future is in jeopardy until it is too late.

Emotions

The teachers' relationships with TAKS test are laden with a wide range of emotions, both positive and negative, often felt almost simultaneously. The teachers reported feeling stress, pressure, anxiety, and tension about how well their students perform in their content area on the Texas State Assessment.

Positive Emotions

Proud. Elation is experienced by teachers when students are successful on the state assessment. "I am so proud when my students pass the exam, especially 'bubble' students. What are 'bubble' students? Bubble students are those who could pass or fail the state assessment by one or two questions."

Neutral Position

No stress. "Personally, I do not feel a lot of stress. I do my job and I do it well. If students fail the TAKS, I will try to do something else the next time. But the system is designed in such a way that some students will not be successful."

Negative Emotions

Frustration. Frustration was the most common negative emotion experienced by teachers.

> Not only are students frustrated, but teachers are, too, not in giving the test, in the money spent, having expected that all students are alike and one treatment fits all. There is a definite relationship between the TAAS and emotions. I think most emotions about the Texas state assessment are negative for teachers and students. As a teacher, I am often nervous or frustrated. I know students have these emotions, as well as fear, and

> sometime I think certain students free hopelessness when discussing the TAAS. There is so much stress associated with the exam!

Failure. According to the focus group, some teachers develop a strong bond with the students in their care and experience the feeling of failure when their students are not successful on the state exam.

> When students fail the state assessment who have honestly worked hard all year, I feel like a failure. I am so disappointed. I have failed a student when he does not pass the "test" for my content area. No matter what you do, sometimes you cannot reach each child. And when that happens I take it in internally . . . year after year. I feel a little hopelessness.

Stress. Negative emotions experienced by students sometimes are presented in physical ways.

> Too much stress! Older students, 9th, 10th, and 11th graders, are very stressed about the state assessment for graduation. Even advanced students suffer from stress and concern about the test. Kids have headaches, ulcers in the mouth; frustration abounds. Not sleeping. Some students are stressed for months and sometimes years before the test. But students are not the only ones who are stressed. Teachers are also stressed, especially Math and English teachers. We all experience some form of negative emotions about the test. Teachers and students suffer from stress and anxiety! Look, there are stress fractures on students and teachers. And there is stress on principals too. They are held accountable and they can lose their jobs. It has not happened here, but there are places where "good" principals may lose their jobs if the school assessment scores are low.

Fear. Teachers consider fear as one of the most disabling emotions experienced by students.

> Many students are poor test-takers. Low-performing students are sometimes so fearful of the test . . . they tend to give up quickly. And when a poor kid in an honors classes fails the state assessment, their embarrassment is so high. Some kids act like they are not concerned about the test but they are. They say things like "yeah, yeah it a big test, but who cares?" Regular students have more anxiety, stress, and frustration. They fear looking stupid. Being labeled stupid. Not going to graduate. Motivation encompassed positive and negative influences on students to obtain an education and to perform at an optimum levels in school.

Motivation.

According to the focus group, the affinity Motivation encompassed positive and negative influences on students to obtain an education and to perform at an optimum academic level in school.

It is hard. Student motivation is one of largest challenges facing teachers.

> Motivating students who have failed the TAAS or who are afraid of failing TAAS is hard. Some students bring so much baggage about state assessment, that by the time they get to high school it is very difficult to motivate them. It requires a lot of energy. “At-risk” students are afraid of failing and they know what happens to people who do not have a high school diploma. It is hard to motivate the students. Some of them do not see a need to take the practice tests. [Laugh] I have a hard time motivating myself sometimes for TAAS practice and review.

Upper-level kids are not interested in classes, because we are not giving them what they need. Some teachers have experienced difficulty motivating students within current instructional guidelines.

> They are not being challenged. If we teach challenging concepts with a high level of expectations in an atmosphere where every student can be successful, then students would be motivated. If not, they will be bored. Also, we need to stop labeling students so early. Students are labeled and these labels stick throughout high school.”

It is a problem. The teachers acknowledged that some students are not aware of the possible negative or unintended consequences high-stakes testing may have on their lives.

> Getting kids to take the test seriously is a problem. They have had enough. Students do not see the use for the test. Smart kids worry too much. And the slowest kids do not realize the problem. Many students are not motivated to learn information that may not be on the state assessment. If it is not on the TAKS, many students do not pay attention in class or feel they need to know the information. Look, it’s hard to motivate kids for more tests. If it is not on the TAKS, often students don’t feel they do not need to know it or pay attention.

Motivating students is an art. Here, capable teachers share positive experiences with student motivational techniques and strategies.

> Motivation of students is my biggest thing. It is an art. How kids feel to a large part determine how well they learn. We have so many wasted kids who are just turned off by school. It is very hard to motivate scared people, but I try. I have found that if I am teaching challenging materials and I explain it in such a way everyone understands, kids are motivated. But if students feel that they are not going to succeed, it is hard to motivate them. Dropouts are those students who leave school because they repeatedly fail the state assessment or they felt that they would not be able to pass the exam.

Dropouts

> This affinity was defined by the participants in Group III as students who leave school because they repeatedly fail the state assessments; students who left school because they felt they would not be able to pass the exam; and/or students who complete all high school class work, fail the state assessment, and quietly leave at the end of their senior year.

Unprepared. The focus group teachers offer their opinion as to why students drop out of school.

> Dropouts are a product of a "sick" system. A number students who dropped out of school were minorities. And I am certain some of these kids dropped out because of state assessment. Loss of self-esteem and loss of hope for their future cause kids to drop out. *I do not think the state assessment causes students to drop out of school*. *They drop out because they are not prepared; we have not prepared them.* We need to do a better job with at-risk students, low achievers, and students with different learning styles. However, the dropout rate would likely decrease if diplomas no longer revolved around the TAAS/TAKS test. Look—dropouts are by-products of a sick system. I do think students drop out of school because they cannot pass the state assessment. We need to ask why we are not preparing students to be successful. It is the system that has failed the student, not the student that has failed the system. We [teachers and/or school] have failed students who trust that we know what we are doing. If we strengthen the system, I think we would have fewer students leaving school."

The Assessment. There was not a consensus why students drop out of school; however, some teachers viewed state assessments as a contributing factor.

> We know that the test caused dropouts, or kids are pushed out before the test. But I bet you are not going to find an administrator in the state that will admit it. Over the years there are some kids I cannot forget, and how they were treated. Our yearly dropout rate is getting better. But if you want the truth, or a better truth, look at the number of kids who enter 9th grade, and the number of students who graduate four years later. Those are the kids we lose . . . dropped out . . . pushed out . . . moved out, or however you want to look at it. Look at the PEIMS data over a four-year period and you will get a different picture.

No effect. Teachers communicated that there was no relationship between the TAKS and school dropout rates under certain conditions.

Personally, *I do not think TAKS affects the dropout rate if students have a good teacher.* In my classes, students see that they are learning and they stay and keep trying, but dropouts could be caused by failing the TAKS."

Bias

Bias describes negative cultural, racial, class-status misconceptions held about students.

Language. The teachers were concerned that language, and not content knowledge, was being evaluated in some test questions.

> The test is unfair to many students. How do you assess higher math skills when reading becomes a factor? Wording on the test may not match the way we teach. Language contained in the state assessment is often difficult for some English as a Second Language learners. And wording on the test does not reflect how students in different regions of the state are taught. The way we say words in the South and what those words mean on the TEKS and TAAS/TAKS do not always mean the same. When students see a word in a question that is unfamiliar, often they freeze and cannot or do not answer the question correctly. The exams have improved and I see less bias but the state needs to do more. Wording on the state

> assessment is geared to middle-class “white” America; those kids who do not fit this mold suffer.

***Socio-economic*.** The teachers communicated that socio-economic status may prohibit students from successfully answering state assessment questions. **“**The bias I associate with TAAS and TAKS is socio-economical. Some students do not come to school with the vocabulary of ‘middle’ America and are at a handicap when taking the test.”

Unaware. Comments below indicate that some teachers may benefit from diversity-focused professional development.

> I do not see a lot of bias. I teach mainly 504, Special Education and minority students. I know that bias exists, but I am not equipped to identify it. However, sometimes people call my attention to biased questions; then I get it. I guess I do not see the bias because the test represents my culture. Over the years I have seen bias. And of course it affects kids. I could easily recognize racial bias. But I am not quite sure if I notice cultural or language bias as easily. This is an area that most teachers need professional development.

Higher Teacher Accountability

The teachers in the study defined this affinity as higher certification requirements before teachers were licensed to teach, including additional coursework in childhood development, longer practice-teaching internships, and diversity training. They also included reward or punishment associated with the number of their students who passed or failed the state assessment, as well as the number professional development hours required maintaining teaching certificates.

The participants in Group III perceived that many teachers who taught in their school were highly qualified and worked effectively with most students in their care. However, they did not want the state of Texas to tie teacher

accountability to the TAKS assessment. This is illustrated in the comments below:

Change and Exposure. The focus group indicated that accountability provides a way for the outstanding skills of many teachers to be exposed to attain community exposure.

> The state has a lot of wonderful teachers and they are doing a better job of presenting a balanced curriculum to all students, including poor and minority students. Many teachers performed outstanding teaching before the state assessment. But there are people in teaching like any other profession who just want a "pay check," but accountability provides an excellent opportunity for a community to realize what outstanding jobs some teachers performed in preparing students.

Teacher accountability should not be tied to state assessment. Teachers were concerned that state assessment may unduly influence teachers' certification policies.

> The state has a lot of wonderful teachers and they are doing a better job of presenting a balanced curriculum to all students, including poor and minority students. I was a good teacher long before the state assessment. I know a lot more about what students need than those Representatives we have down there in Austin. Teachers should be accountable. And those teachers who are not suitable should leave the classroom. But teacher accountability should not be tied to the state assessment.

Class Size/Resources

Teachers described the affinity *Class Size/Resources* as the number of students assigned to a class. Resources included class materials and technology used to enhance learning.

Smaller Classes. Teachers voiced opinions that large classes and inadequate resources negatively affect student learning and instruction.

> It is outrageous to have thirty students in a regular class. We need smaller class sizes and more materials to teach "At Risk," 504, and Special Education students. Sometime we are flying blind. We need tools, such as

> computer programs, to help assess the needs of students and provide the educational interventions they need. It is outrageous to have thirty students in a regular class. Thirty AP kids, fine. But regular kids, instead of twenty, where you have more one-to-one contact. You can look at learning styles, see what they need, and better equip each student to be successful. Partner up kids, meet their needs. It is impossible to address the needs of 30 students . . . when, like I have in one of my classes, 504 students, six Special Ed, two ESL, 20 other kids of which two have been kicked out of their house and one girl is upset because of a nasty breakup with a boyfriend. To reach kids, we need more time, resources, and smaller classes. We need smaller classes for regular and advanced students. It is almost impossible to give each student the type of attention needed in large classes. We also need additional support personnel for "low achievers" as well as 504 and special-needs children.

Class materials. Some teachers indicated they experience inadequate classroom resources.

> We need class materials, especially for science. Class materials aligned to the TEKS. Resources—we need many, many, many more. You get what you pay for. We need equipment, books that address the TAKS. And we have too many students in classes. And we need to use class time like it is a true resource. As well as state TEKS that directly address the TAKS. *But where will the money come from? How can the district pay for everything?*

Group Reality: System Relationships

Overview

IQA methodology allows researchers to draw a picture of a given system, a System Influence Diagram (SID), reflecting a theory of perception held by a group—its group reality. This reality is based on system relationships, which are refined using theoretical coding. Theoretical coding refers to the cause-and-effects relationships among the affinities or elements in the system. The SID is produced using a set of rules for rationalization based on a summary of theoretical codes called an Interrelationship Diagram (IRD) for a group or individual (Northcutt & McCoy, 2000).

All possible direct links between the affinities are investigated and hypotheses are developed using data. In this study, this hypothetical construction takes the form of an if . . . then statement. Members of the focus groups were asked to determine the relationship between all possible pairs of affinities. For any two affinities there are only three relationships: A influences B; B directly influences A; or there is no relationship between A and B. Hypotheses are recorded on a protocol called the Affinity Relationship Table (ART). This begins with the second part of the interview, called the theoretical interview.

Interview Protocol Part Two

The structured theoretical interview is designed to identify relationships between affinities. It is presented through an Affinity Relationship Table. The table provides a quick reference of all of the possible relationships between affinities. A sample of an interview protocol used with Group I participants during the interview, including an Affinity Relationship Table, follows:

Table 4.1
Theoretical Coding
Student Group I

Themes or affinities identified have some kind of relationship; one effects or causes the other. Please look at each theme and decide if or how it relates to each other theme. Tell me about your experiences with such relationships. Please give specific examples of how the relationships have affected your experience.

Affinity Name
1. All or Nothing
2. Emotions/Stress
3. Dropouts
4. More time to learn required material
5. Practice tests
6. Material tested
7. Use of class time

Possible Relationships
A → B
A ← B
A <> B (No Relationship)

Group I – Students Who Failed the TAKS
Affinity Relationship Table

Affinity Pair Relationship	Affinity Pair Relationship	Affinity Pair Relationship
1 2	2 4	3 7
1 3	2 5	4 5
1 4	2 6	4 6
1 5	2 7	4 7
1 6	3 4	5 6
1 7	3 5	5 7
2 3	3 6	6 7

Interview protocols for Groups II and III are included in Appendices C and D. Frequency tabulations of the Affinity Pair Relationships developed from the Theoretical Interview Protocols, and Affinity Pair tables for each group, are presented following the next section for each focus group.

Transcripts and Theoretical Code Tables

After the interviews had been transcribed, the text was analyzed for theoretical codes which illustrate a relationship between two or more affinities. The reference was then documented for retrieval by recording the affinity number on the line of transcript that refers to the affinity Individual Interview Theoretical Code Table (TCT). Quotes relating to a specific affinity pair relationship were cut and pasted into the third column of the TCT, along with the line(s) of the transcript that were the source of the theoretical quote. Once coding of the interviews was completed, the data were summarized to create a composite of the individuals' experience with the phenomenon. Theoretical data were transferred from each Individual Interview Theoretical Code Table to a Combined Interview Theoretical Code Table, thus creating a database for the entire set of respondents containing all theoretical codes for all affinity pairs, with each code containing a link or a reference to the transcript and line numbers that produced the code.

Theoretical Code Frequency Table

Next, the Combined Interview Theoretical Code Affinity Relationship Table is created. Individual participants may disagree about the direction of a relationship; the table lists both directions for relationships. The researcher counted the number of respondents who identified the relationship in the same direction and place the tally in the frequency. The same was done for all participants who identified the relationship in the opposite direction. A separate Theoretical Code Frequency Table was created for each group. Below are the Theoretical Code Frequency Tables for all three groups.

Table 4.2
Group I – Combined Interview
Theoretical Code
Frequency Table
(Students Who Have Failed the TAAS/TAKS)

Affinity Pair Relationship	Frequency	Affinity Pair Relationship	Frequency	Affinity Pair Relationship	Frequency
1 → 2	8	2 → 4	0	3 → 7	1
1 ← 2	0	2 ← 4	0	3 ← 7	7
1 → 3	8	2 → 5	2	4 → 5	3
1 ← 3	0	2 ← 5	6	4 ← 5	5
1 → 4	5	2 → 6	1	4 → 6	2
1 ← 4	3	2 ← 6	7	4 ← 6	6
1 → 5	6	2 → 7	0	4 → 7	6
1 ← 5	2	2 ← 7	0	4 ← 7	2
1 → 6	5	3 → 4	0	5 → 6	0
1 ← 6	3	3 ← 4	8	5 ← 6	0
1 → 7	7	3 → 5	1	5 → 7	7
1 ← 7	1	3 ← 5	7	5 ← 7	1
2 → 3	7	3 → 6	2	6 → 7	7
2 ← 3	1	3 ← 6	6	6 ← 7	1

Table 4.3
Group II – Combined Interview
Theoretical Code
Frequency Table
(Students Who Passed the TAAS/TAKS)

Affinity Pair Relationship	Frequency	Affinity Pair Relationship	Frequency	Affinity Pair Relationship	Frequency
1 → 2	3	2 → 4	7	3 → 7	8
1 ← 2	5	2 ← 4	1	3 ← 7	0
1 → 3	0	2 → 5	5	4 → 5	3
1 ← 3	8	2 ← 5	3	4 ← 5	5
1 → 4	1	2 → 6	8	4 → 6	5
1 ← 4	7	2 ← 6	0	4 ← 6	3
1 → 5	3	2 → 7	6	4 → 7	6
1 ← 5	5	2 ← 7	2	4 ← 7	2
1 → 6	1	3 → 4	2	5 → 6	4
1 ← 6	7	3 ← 4	6	5 ← 6	4
1 → 7	7	3 → 5	2	5 → 7	7
1 ← 7	1	3 ← 5	6	5 ← 7	1
2 → 3	8	3 → 6	3	6 → 7	8
2 ← 3	0	3 ← 6	5	6 ← 7	0

Table 4.4
Group III – Combined Interview
Theoretical Code
Frequency Table
(High School Teachers)

Affinity Pair Relationship	Frequency	Affinity Pair Relationship	Frequency	Affinity Pair Relationship	Frequency
1 → 2	9	2 → 8	2	5 → 6	8
1 ← 2	0	2 ← 8	7	5 ← 6	1
1 → 3	9	2 → 9	7	5 → 7	4
1 ← 3	0	2 ← 9	2	5 ← 7	5
1 → 4	9	3 → 4	8	5 → 8	0
1 ← 4	0	3 ← 4	1	5 ← 8	0
1 → 5	8	3 → 5	7	5 → 9	1
1 ← 5	1	3 ← 5	2	5 ← 9	8
1 → 6	8	3 → 6	8	6 → 7	4
1 ← 6	1	3 ← 6	1	6 ← 7	5
1 → 7	7	3 → 7	6	6 → 8	2
1 ← 7	2	3 ← 7	3	6 ← 8	7
1 → 8	8	3 → 8	2	6 → 9	0
1 ← 8	1	3 ← 8	7	6 ← 9	9
1 → 9	9	3 → 9	0	7 → 8	0
1 ← 9	0	3 ← 9	0	7 ← 8	0
2 → 3	0	4 → 5	4	7 → 9	0
2 ← 3	9	4 ← 5	5	7 ← 9	0
2 → 4	7	4 → 6	8	8 → 9	3
2 ← 4	2	4 ← 6	1	8 ← 9	6
2 → 5	7	4 → 7	4		
2 ← 5	2	4 ← 7	5		
2 → 6	7	4 → 8	3		
2 ← 6	2	4 ← 8	6		
2 → 7	0	4 → 9	3		
2 ← 7	0	4 ← 9	6		

Pareto Protocol

The Pareto Protocol is a statistical method used to allow the researcher to represent the consensus or the preponderance of the group analysis of

relationships. A Pareto chart is a special kind of bar chart and is used to display the relative importance of a problem or condition. IQA uses the Pareto Principle operationally deployed to achieve and document a degree of consensus in a focus group (Northcutt & McCoy). The results of the frequency tallies were transferred into the Pareto Protocol Tables. The Pareto Protocol determined which affinity pair relationships were used in the system. The Pareto Tables for Groups I, II, and III presented below:

Table 4.5
Group I – Affinities in Descending Order of Frequency With Pareto and Power Analysis (Students Who Have Failed the TAAS/TAKS)

Affinity Pair Relationship	Frequency	Cumulative Frequency	Cumulative Percent (Relation)	Cumulative Percent (Frequency)	Power
1. 1 > 2	8	8	3.8	5.8	2.0
2. 1 > 3	8	16	7.7	11.7	4.0
3. 3 < 4	8	24	11.5	17.5	6.0
4. 1 > 7	7	31	15.4	22.6	7.2
5. 2 < 6	7	38	19.2	27.7	8.5
6. 2 > 3	7	45	23.1	32.8	9.8
7. 3 < 5	7	52	26.9	38.0	11.0
8. 3 < 7	7	59	30.8	43.1	12.3
9. 5 > 7	7	66	34.6	48.2	13.6
10. 6 > 7	7	73	38.5	53.3	14.8
11. 1 > 5	6	79	42.3	57.7	15.4
12. 2 < 5	6	85	46.2	62.0	15.9
13. 3 < 6	6	91	50.0	66.4	16.4
14. 4 < 6	6	97	53.8	70.8	17.0
15. 4 > 7	6	103	57.7	75.2	17.5
16. 1 > 4	5	108	61.5	78.8	17.3
17. 1 > 6	5	113	65.4	82.5	17.1
18. 4 < 5	5	118	69.2	86.1	16.9
19. 1 < 4	3	121	73.1	88.3	15.2
20. 1 < 6	3	124	76.9	90.5	13.6
21. 4 > 5	3	127	80.8	92.7	11.9
22. 1 < 5	2	129	84.6	94.2	9.5
23. 2 > 5	2	131	88.5	95.6	7.2
24. 3 > 6	2	133	92.3	97.1	4.8
25. 4 < 7	2	135	96.2	98.5	2.4
26. 4 > 6	2	137	100.0	100.0	0.0
Total Frequency	**137**	**Equal Total Frequency**	**Equals 100%**	**Equals 100%**	**Power = E-D**

Table 4.6
Group II – Affinities in Descending Order of Frequency With Pareto and Power Analysis
(Students Who Have Passed the TAAS/TAKS)

Affinity Pair Relationship	Frequency	Cumulative Frequency	Cumulative Percent (Relation)	Cumulative Percent (Frequency)	Power
1. 1 < 3	8	8	3.6	5.2	1.6
2. 2 > 3	8	16	7.1	10.3	3.2
3. 2 > 6	8	24	10.7	15.5	4.8
4. 3 > 7	8	32	14.3	20.6	6.4
5. 6 > 7	8	40	17.9	25.8	7.9
6. 1 < 4	7	47	21.4	30.3	8.9
7. 1 < 6	7	54	25	34.8	9.8
8. 1 > 7	7	61	28.6	39.4	10.8
9. 2 > 4	7	68	32.1	43.9	11.7
10. 5 > 7	7	75	35.7	48.4	12.7
11. 2 > 7	6	81	39.3	52.3	13
12. 3 < 4	6	87	42.9	56.1	13.3
13. 3 < 5	6	93	46.4	60	13.6
14. 4 > 7	6	99	50	63.9	13.9
15. 1 < 2	5	104	53.6	67.1	13.5
16. 1 < 5	5	109	57.1	70.3	13.2
17. 2 > 5	5	114	60.7	73.5	12.8
18. 3 < 6	5	119	64.3	76.8	12.5
19. 4 < 5	5	124	67.9	80	12.1
20. 4 > 6	5	129	71.4	83.2	11.8
21. 5 < 6	4	133	75	85.8	10.8
22. 5 > 6	4	137	78.6	88.4	9.8
23. 1 > 2	3	140	82.1	90.3	8.2
24. 1 > 5	3	143	85.7	92.3	6.5
25. 2 < 5	3	146	89.3	94.2	4.9
26. 3 > 6	3	149	92.9	96.1	3.3
27. 4 < 6	3	152	96.4	98.1	1.6
28. 4 > 5	3	155	100	100	0

Total Frequency	155	Equal Total Frequency	Equals 100%	Equals 100%	Power = E-D

Table 4.7
Group III – Affinities in Descending Order of Frequency With Pareto and Power Analysis (High School Teachers)

Affinity Pair Relationship	Frequency Sorted (Descending)	Cumulative Frequency	Cumulative Percent (Relation)	Cumulative Percent (Frequency)	Power
1. 1 > 2	9	9	2.6	3.6	1
2. 1 > 3	9	18	5.1	7.1	2
3. 1 > 4	9	27	7.7	10.7	3
4. 1 > 9	9	36	10.3	14.2	4
5. 2 < 3	9	45	12.8	17.8	5
6. 6 < 9	9	54	15.4	21.3	6
7. 1 > 5	8	62	17.9	24.5	6.6
8. 1 > 6	8	70	20.5	27.7	7.2
9. 1 > 8	8	78	23.1	30.8	7.8
10. 3 > 4	8	86	25.6	34	8.4
11. 3 > 6	8	94	28.2	37.2	8.9
12. 4 > 6	8	102	30.8	40.3	9.5
13. 5 < 9	8	110	33.3	43.5	10.1
14. 5 > 6	8	118	35.9	46.6	10.7
15. 1 > 7	7	125	38.5	49.4	10.9
16. 2 < 8	7	132	41	52.2	11.1
17. 2 > 4	7	139	43.6	54.9	11.4
18. 2 > 5	7	146	46.2	57.7	11.6
19. 2 > 6	7	153	48.7	60.5	11.8
20. 2 > 9	7	160	51.3	63.2	12
21. 3 < 8	7	167	53.8	66	12.2
22. 3 > 5	7	174	56.4	68.8	12.4
23. 6 < 8	7	181	59	71.5	12.6
24. 3 > 7	6	187	61.5	73.9	12.4
25. 4 < 8	6	193	64.1	76.3	12.2
26. 4 < 9	6	199	66.7	78.7	12
27. 8 < 9	6	205	69.2	81	11.8
28. 4 < 5	5	210	71.8	83	11.2

Table 4.7 (Continued) Group III – Affinities in Descending Order of Frequency With Pareto and Power Analysis (High School Teachers)					
Affinity Pair Relationship	**Frequency Sorted (Descending)**	**Cumulative Frequency**	**Cumulative Percent (Relation)**	**Cumulative Percent (Frequency)**	**Power**
29. 4 < 7	5	215	74.4	85	10.6
30. 5 < 7	5	220	76.9	87	10
31. 6 < 7	5	225	79.5	88.9	9.4
32. 4 > 5	4	229	82.1	90.5	8.5
33. 4 > 7	4	233	84.6	92.1	7.5
34. 5 > 7	4	237	87.2	93.7	6.5
35. 6 > 7	4	241	89.7	95.3	5.5
36. 3 < 7	3	244	92.3	96.4	4.1
37. 4 > 8	3	247	94.9	97.6	2.8
38. 4 > 9	3	250	97.4	98.8	1.4
39. 8 > 9	3	253	100	100	0
Total Frequency	**253**	**Equal Total Frequency**	**Equals 100%**	**Equals 100%**	**Power = E-D**

Affinity Relationship Table

The Affinity Relationship Table (ART) summarizes the relationships to be used in the systems as indicated by the Pareto Protocol. In Group I (students who have failed the TAAS/TAKS assessment), since there are seven affinities, there are twenty-one different possible pairs for which a cause and relation exists. Group II (students who have passed the TAAS/TAKS assessment) also contains 21 affinity pairs. And Group III (secondary teachers) has 56 affinity pairs.

In Group I, of the 21 relationships, eight were considered ambiguous or in conflict; Group II contained seven conflicting relationships and Group III had 8 ambiguous relationships.

Imagine that a group has written a number of hypotheses arguing that affinity A influences affinity B (A→B). And another set of hypotheses argues the opposite, that B→A. The argument will not be resolved with a submission to the Pareto Chart. The table will contain hypotheses that argue for both directions, and both sets seem equally credible. These conflicting relationships are an indication that a feedback loop may be present in the system; if so, it will be addressed later. Now the researcher discards the lowest-frequency conflict and uses the highest-frequency affinity pair to build the system. For the ARTs for the three groups and conflicting relationships tables see Appendix E.

Interrelationship Diagram (IRD)

As discussed in Chapter III, an Interrelationship Diagram (IRD) is a matrix containing all the perceived relationships in the system, classifying the affinities as drivers or outcomes. This activity, theoretical coding, creates an extended reality for the group through further discourse. Creating IRD is the first step in a general process called *rationalizing the system*. Using a forced directional choice in a specific order, focus group participants evaluated if there was a direct cause/effect relationship or if no relationship existed between affinities, creating a skeleton of a theory of perception.

The IRD matrix depicts the relationship or lack of relationships among the affinities. The value of Δ measures the number of times a particular affinity was referred to as the cause or an effect. IRDs created in regard to Focus Groups I, II, and III are shown below.

Table 4.14
Group I – Tabular IRD
(Students Who Failed the TAKS)

	1	2	3	4	5	6	7	OUT	IN	Δ
1		↑	↑	↑	↑	↑	↑	6	0	6
2	←		↑		←	←		1	3	-2
3	←	←		←	←	←	←	0	6	-6
4	←		↑		←	←	↑	2	3	-1
5	←	↑	↑	↑			↑	4	1	3
6	←	↑	↑	↑			↑	4	1	3
7	←		↑	←	←	←		1	4	-3

Table 4.15
Group I – Tabular IRD
Sorted in Descending Order of Δ
(Students Who Failed the TAKS)

	1	2	3	4	5	6	7	OUT	IN	Δ
1		↑	↑	↑	↑	↑	↑	6	0	6
5	←	↑	↑	↑			↑	4	1	3
6	←	↑	↑	↑			↑	4	1	3
4	←		↑		←	←	↑	2	3	-1
2	←		↑		←	←		1	3	-2
7	←		↑	←	←	←		1	4	-3
3	←	←		←	←	←	←	0	6	-6

Count the number of up arrows (↑) or *Outs*
Count the number of left arrows (←) or *Ins*
Subtract the number of *Ins* from the *Outs* to determine the (Δ) *Deltas*
Δ = Out-In

Table 4.16
Group II – Tabular IRD
(Students Who Passed the TAAS/TAKS)

	1	2	3	4	5	6	7	OUT	IN	Δ
1		←	←	←	←	←	↑	1	5	-4
2	↑		↑	↑	↑	↑	↑	6	0	6
3	↑	←		←	←	←	↑	2	4	-2
4	↑	←	↑		←	↑	↑	4	2	2
5	↑	←	↑	↑		↑	↑	5	1	4
6	↑	←	↑	←	←		↑	3	3	0
7	←	←	←	←	←	←		0	6	-6

Table 4.17
Group II – Tabular IRD
Sorted in Descending Order of Δ
(Students Who Passed the TAAS/TAKS)

	1	2	3	4	5	6	7	OUT	IN	Δ
2	↑		↑	↑	↑	↑	↑	6	0	6
5	↑	←	↑	↑		↑	↑	5	1	4
4	↑	←	↑		←	↑	↑	4	2	2
6	↑	←	↑	←	←		↑	3	3	0
3	↑	←		←	←	←	↑	2	4	-2
1		←	←	←	←	←	↑	1	5	-4
7	←	←	←	←	←	←		0	6	-6

Count the number of up arrows (↑) or *Outs*
Count the number of left arrows (←) or *Ins*
Subtract the number of *Ins* from the *Outs* to determine the (Δ) *Deltas*
Δ = Out-In

Table 4.18
Group III – Tabular IRD
(Teachers)

	1	2	3	4	5	6	7	8	9	OUT	IN	Δ
1		↑	↑	↑	↑	↑	↑	↑	↑	8	0	8
2	←		←	↑	↑	↑		←	↑	4	3	1
3	←	↑		↑	↑	↑	↑	←		5	2	3
4	←	←	←		←	↑	←	←	←	1	7	-6
5	←	←	←	↑		↑	←		←	2	5	-3
6	←	←	←	←	←		←	←	←	0	8	-8
7	←		←	↑	↑	↑				3	2	1
8	←	↑	↑	↑		↑			←	4	2	2
9	←	←		↑	↑	↑		↑		4	2	2

Table 4.19
Group III – Tabular IRD – Sorted in Descending Order of Δ
Teachers

	1	2	3	4	5	6	7	8	9	OUT	IN	Δ
1		↑	↑	↑	↑	↑	↑	↑	↑	8	0	8
3	←	↑		↑	↑	↑	↑	←		5	2	3
8	←	↑	↑	↑		↑			←	4	2	2
9	←	←		↑	↑	↑		↑		4	2	2
2	←		←	↑	↑	↑		←	↑	4	3	1
7	←		←	↑	↑	↑				3	2	1
5	←	←	←	↑		↑	←		←	2	5	-3
4	←	←	←		←	↑	←	←	←	1	7	-6
6	←	←	←	←	←		←	←	←	0	8	-8

Count the number of up arrows (↑) or *Outs*
Count the number of left arrows (←) or *Ins*
Subtract the number of *Ins* from the *Outs* to determine the (Δ) *Deltas*
Δ = Out-In

Tentative SID Assignments Table

The relative position of an affinity within the system is marked by the delta value. Affinities with a positive delta are *drivers*, representing the *cause.* And affinities with negative delta values are outcomes, illustrating the effect. As stated in Chapter III, the Tentative SID Assignments Table represents the initial placement of affinities for the system, known as the System Influence Diagram (SID). Below are The Tentative SID Tables for each group.

Table 4.20
Group I – Tentative SID Assignments
(Students Who Failed the TAKS)

1	All or Nothing	Primary Driver
5	Practice Tests	Secondary Driver
6	Materials Tested	Secondary Driver
4	More Time to Learn Required Materials	Secondary Outcome
2	Emotions	Secondary Outcome
7	Used of Class Time	Secondary Outcome
3	Dropouts	Primary Outcome

Table 4.21
Group II – Tentative SID Assignments
(Students Who Passed the TAAS/TAKS)

2	Too Much Weight on the Test	Primary Driver
5	Equity	Secondary Driver
4	Benefits	Secondary Outcomes
6	What I Want to Learn	Secondary Outcome
3	Emotions	Secondary Outcome
1	Motivation	Secondary Outcome
7	Dropouts	Primary Outcome

Table 4.22 Group III – Tentative SID Assignments (Teachers)		
1	Assessment is Weighted Too High	Primary Driver
3	Practice Tests	Secondary Driver
9	Class Size/ Resources	Secondary Driver
8	Teacher Accountability	Secondary Driver
2	Instructional Focus	Secondary Driver
7	Bias	Secondary Driver
5	Motivations	Secondary Outcome
4	Emotions	Secondary Outcome
6	Dropouts	Primary Outcome

System Influence Diagram (SID)

Next, in the process of analyzing the data, the System Influence Diagram (SID), or "mind map," a visual representation of an entire system of influences and outcomes, was developed. The affinities are arranged according to the Tentative SID Assignment chart, creating a flow chart or "mind map" using Inspiration software. Through the use of the computer, the affinities were place in rough order of topological zones: the Primary Drivers to the left of the screen, and the Primary Outcomes to the right. Secondary Drivers and Secondary Outcomes were then placed between the primaries. Each affinity name and number was placed in a square text box. Next, arrows were drawn between affinities in directions of the relationships as represented in the IRD, creating a composite interview uncluttered SID. These are shown, for each group, in the analysis in Chapter V.

Pareto Reconciled SID

Once all redundant links are removed, the Pareto Protocol was examined for conflicting relationships. A conflicting relationship occurs when the same affinity pair has relationships in both directions and a significant frequency to include both in the system. Earlier, the lesser-frequency was temporarily ignored in the IRD, but it is reconciled in the uncluttered SID. To account for the relationships, the system was examined to see if the conflicting relationship was indicated in the system, possibly as part of a feedback loop. If such was the case, nothing needed to be done. An arrow was placed from Comprehension to IQA as a Process to reconcile one of these conflicts.

The Composite Interview Uncluttered SID

The cluttered SID contains all of the relationships described by the group. It is saturated with relationships that are interlocked with the system. It is very comprehensive and laden with relationships, containing many redundant links. This makes it very difficult to interpret a system even with a few affinities. A way to reconcile the system is to remove redundant links, producing an uncluttered composite SID (theory of perception).

In Chapter V, the system for each group will be presented, showing several versions: (1) a cluttered SID, which is first version of the SID containing each link presented in the IRD; (2) the Pareto reconciliation process; and finally, (3) the final system or uncluttered SID (theory of perception).

CHAPTER V – ANALYSIS OF DATA FINDINGS

"A vision without action is merely a dream. Action without a vision just passes time. Vision with action can change the world"
Joel Baker (Bowsher, 2001, p. 259)

Introduction

It is not necessary to change the world, but, as stated in Chapter II, there is a critical challenge to change the American public school system, so it meets the needs of all children. The findings contained in Chapter IV allow the researcher and others to readily analyze consequences of high-stakes testing from the critical perspectives of students and teachers, focusing on the following research questions:

- What critical perspectives do teachers (and students) have about the effects of high-stakes testing?
- How do intended and unintended consequences of high-stakes testing affect students?
- How do intended and unintended consequences of high-stakes testing affect instruction and classroom practices?

The Interactive Qualitative Analysis method was used to collect the data. Groups participating in the study identified affinities. The researcher conducted axial and theoretical coding. The group Interrelationship Diagram (IRD) summarized the results of the group theoretical coding. Finally, by taking the information found in the IRD and presenting visual representation of affinities and relationships among affinities, a System Influence Diagram (SID) was created. In

this chapter, additional analysis of the SIDs will be provided, and the systems will be compared.

Touring The Systems

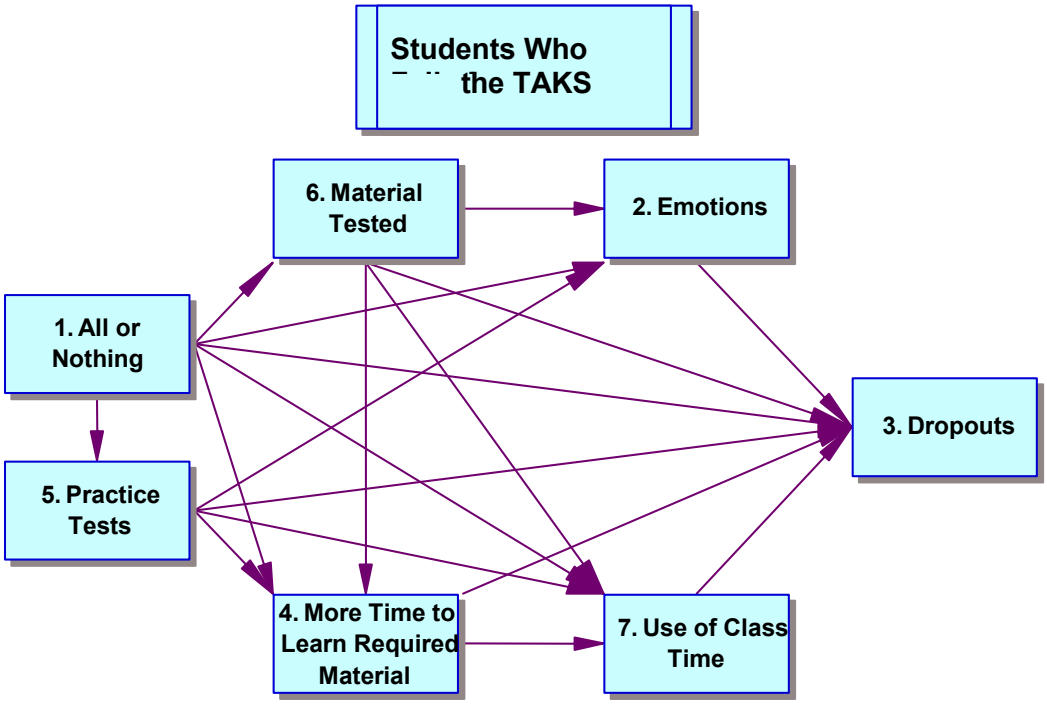
Group I – Analysis Process

The System Influence Diagram (SID) produced by Focus Group I (students who failed the TAAS/TAKS exam) showed several patterns and revealed how students perceive the Texas State Graduation Assessments Policy and consequences of that policy. As previously related, students were asked, “Tell me about your experiences with Texas State Assessment Exams, commonly called the TAAS/TAKS.” Students chose the following affinities or themes:

Figure 5.1
Group I – Affinity Choices

The first version of the SID contains each link present in the IRD. The system is cluttered or saturated with links, producing a rich and comprehensive system. The cluttered SID for Group I is shown below:

Figure 5.2
Group I – Cluttered SID



Parsimony is required. The system has so many links that the explanatory power of the system becomes entangled in details of relationships. This richness must be reconciled by removing redundant links in the system producing a secondary SID called an uncluttered SID. But first, conflicting relationships must be resolved.

Redundant links are removed according to their delta and SID assignment. The relationship between the highest positive delta and the highest negative delta is examined. If there is a path between the two deltas other than a direct link, the link is removed and the process is repeated until all redundancy is eliminated.

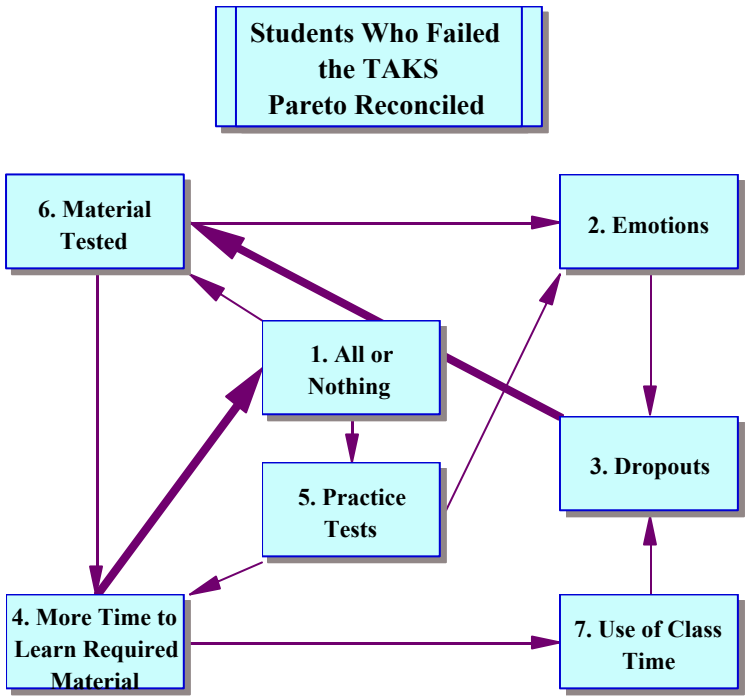
Once all redundant links are removed, the Pareto Protocol is examined for conflicting relationships. A conflicting relationship occurs when the same affinity pair has relationships in both directions, and a significant frequency to include both in the system. In Group I, conflicting relationships existed as illustrated in the table below:

Table 5.1
Group I – Conflicting Relationships

Affinity Pair Relationship		Affinity Pair Relationship		Affinity Pair Relationship
$1 \leftarrow 4$		$2 \rightarrow 5$		$4 \rightarrow 5$
$1 \leftarrow 5$		$3 \rightarrow 6$		$4 \rightarrow 6$
$1 \leftarrow 6$		$4 \leftarrow 7$		

Earlier, the lesser frequencies were temporarily ignored in the IRD, but they are reconciled here in the uncluttered SID. To account for the relationships, the system is examined to see if the conflicting relationship is indicated in the system, possibly as part of a feedback loop. Recursion or feedback loops exist. A feedback loop requires at least three affinities and has no end or no beginning. It is like a circle; the previous affinity influences the successive one, which in turn influences pervious affinities. Recursion is illustrated in following Pareto chart.

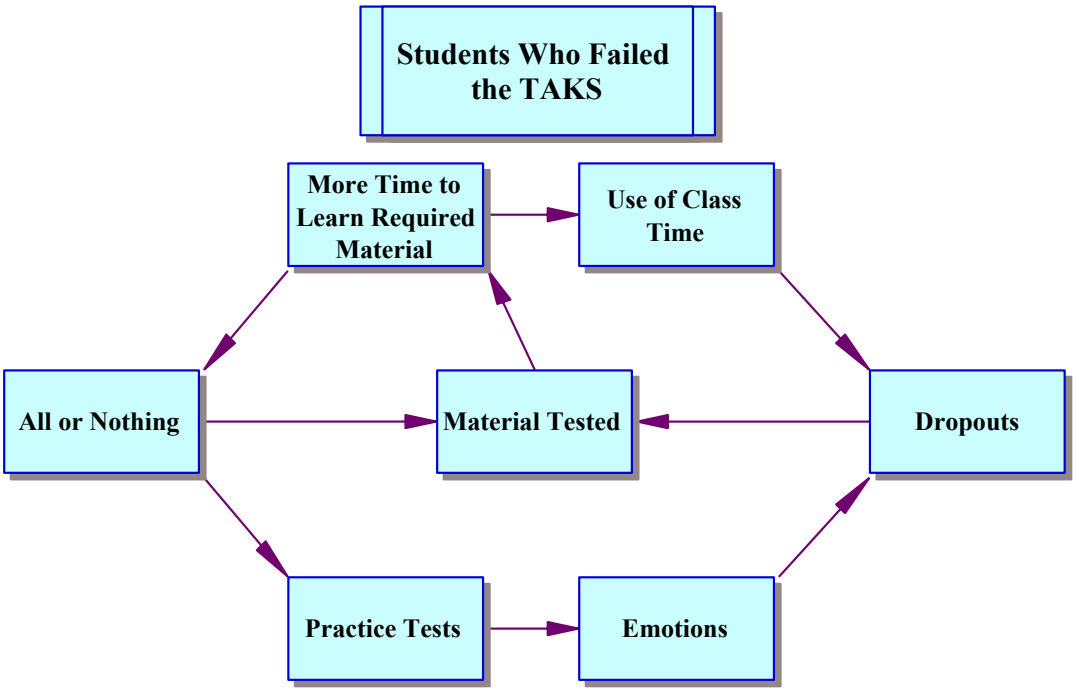
Figure 5.3
Group I – Pareto Reconciled SID



IQA protocol allows the identification of recursion. This is a unique quality of the methodology that distinguishes it from traditional or quantitative path analytic approaches (Northcutt & McCoy, 2000). Next, the System of Influence Diagram (SID) is revealed. A SID may be considered as a set of qualitative structural equations or as a path diagram, where recursion or feedback loops are allowed (Northcutt & McCoy, 2000).

The comments, views, and experiences of the students in Group I, solidified via interviews, were used to illustrate a theory of perception (i.e., the ultimate, uncluttered SID). The following is a visual representation of that system.

Figure 5.4
Group I – Theoretical Perceptions



Group I perceived the Texas State mandated graduation (TAAS/TAKS) assessment, requiring students to pass the graduate exit exam before receiving a high school diploma, as the main driving force of the system. Failure to pass the assessment led students to conclude that their educational future was not very promising, often ending their dreams for college or a secure financial future; hence they labeled this affinity *All or Nothing*.

The students commented that one test should not determine if a student graduates from high school. They expressed concern that there was too much value or weight placed on the Texas high school graduation exam. Students surmised that they and many other students work very hard day-in and day-out to complete the requirements for each high school course. From Group One's viewpoint, the results of twelve or more years of school should be represented by

more than one multiple-choice test. Students remarked that assessments like the TAAS/TAKS tests do not show what students know or what students have been taught. "*I don't think the test should determine everything. At the end what it comes down to is if you pass the test or not. If you don't, then all of the work you have done has been wasted. I know a lot more than what was on that test."* From the viewpoint of most students in this focus group, schooling should represent more than the results from one test. And they suggest a different form of assessment such as a portfolio.

According to Ancess and Witherle (2002), Portfolio-Based Performance Assessment provides an intellectual framework for students to demonstrate in multiple, complex ways that they have met standards for graduation. It provides students with intense, highly personalized teacher support and guidance and allows students to build ownership and confidence in their intellectual capital by integrating multiple opportunities for self-correction and improvement.

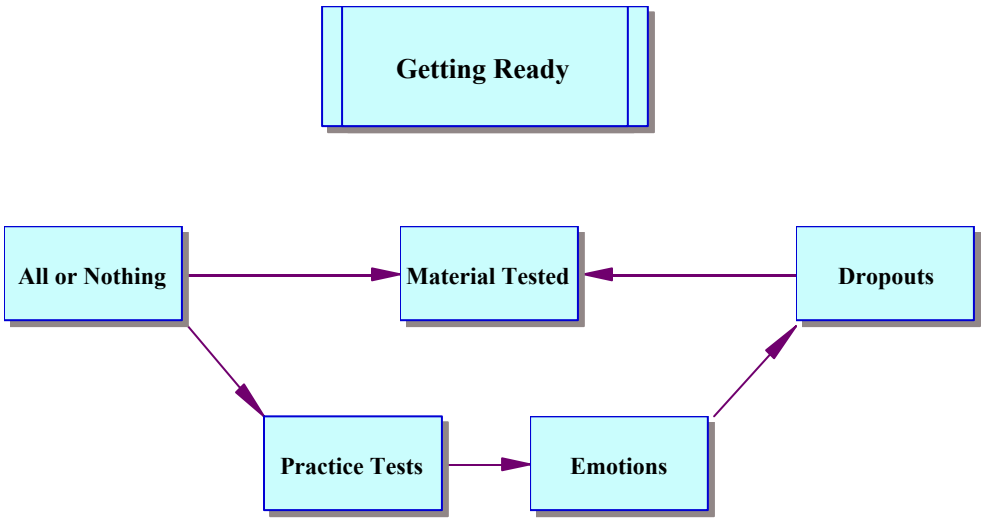
The theory of perception revealed by the students who have failed the TAAS/TAKS comprises one main loop, containing three smaller loops. This main loop is labeled **Try Anyway**. It includes these affinities: *All or Nothing*, *Practice Tests*, *Emotions*, *Dropouts, Use of Class Time*, and *More Time to Learn Required Materials*. Although students have failed the Texas state graduation assessment they continue to pursue their dream of acquiring a high school diploma, and hope their second try to pass the assessments will be successful.

Students were keenly aware of the consequences of failing the assessment a second time, as illustrated by this comment: "I do not think one test should destroy someone's future. What do they think we are going to do? Without a high school diploma you cannot make a honest living. So I just try to

learn what I can and hope I pass the TAAS/TAKS." This is a critical observation and question for policymakers. Hauser (1999) reminds us that young people who fail to get a high school diploma also have lower ratings on quality-of-life factors, starting and maintaining a family, participating in civic activities, and maintaining good health.

The next section of the system, **Getting Ready,** examines the affinities *Practice Tests, Emotions, Material Tested*, and *Dropouts*.

Figure 5.5
Group I – Getting Ready



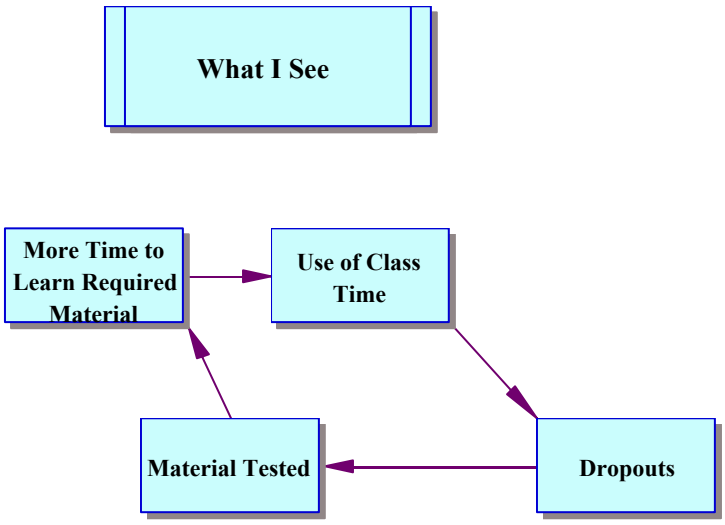
Here, students make the critical decision to drop out or remain in school. This decision is often based on their emotional outlook and how well they perceive their mastery of TAKS-required materials. Many emotions reported by this group were negative. Students' emotions ranged from fear of taking the exams, to family disgrace, to hopelessness and possible lost dreams for better lives. Students who continued to fail the practice tests or the TAKS exams after years of review and study were often discouraged, disappointed, and

disillusioned, and they either dropped out of school or merely tended to leave school after completing the state-required courses without acquiring a high school diploma. Nevertheless, the students in this focus group all hoped they would be successful and pass the state assessment.

Materials tested on the state assessment also dictated the use of practice tests as well as the quantity of practice tests. Practice tests were often seen by the students as a way teachers presented and reviewed disassociated information for the state exam. For example, one students recalled, *"Many times we were only studying sample items for the TAAS/TAKS. Sometimes I am not quite sure how some questions related to the course I am taking. I just try to learn what I can and hope I pass the TAAS/TAKS."* Others students stated that they tried to memorize the questions and answers. Students also remarked that some of the questions on the practice exams for TAKS had very little to do with the their current course of study. Students also inquired why they had encountered specific questions on materials that were last taught to them in middle school. However, students felt they were becoming better test-takers.

The next loop is called **What I See.** This is a feedback loop, containing the affinities *Materials Tested*, *Use of Class Time*, and *Dropouts*.

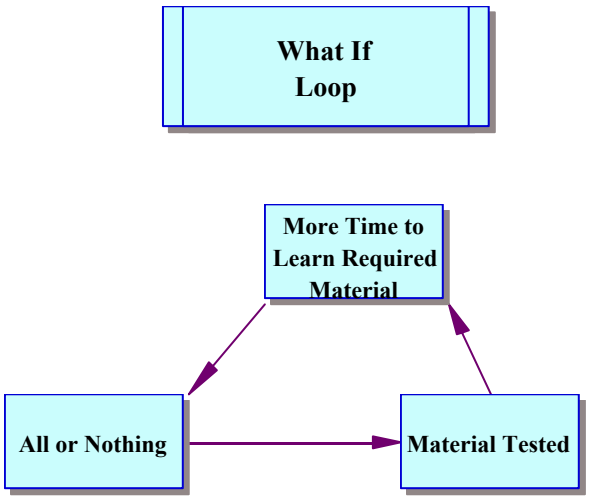
Figure 5.6
Group I – What I See Loop



Students perceived that the use of class time and practice tests used to prepare them for the Texas state assessment were not always productive or successful, creating or fostering negative emotions such as fear and/or frustration. They reported that a large amount of their class time was used to prepare for the TAKS test via review and worksheets, without linking concepts in a meaningful way that they could understand. They felt that their time in class should be used for more than an exercise to pass the state assessment.

The last feedback loop, named **What If,** contains the following themes or affinities: *Materials Tested*, *More Time to Learn Required Materials*, and *All or Nothing*.

Figure 5.7
Group I – What If Loop



Here, students in the study voiced their views that if "slow students" or "students with different learning styles" were given additional time to learn, they could master the required concepts mandated by the state. One student stated, *"There should be more time for students to learn who are not fast learners, 'slow' or 'different' learners*." When asked to define "different" learners, the students remarked that different learners were students who saw information in different patterns or systems than most students. They were the students who were not typical "A" or "B" students. These are students whom teachers do not consider to be very intelligent, because they just learn things in different ways.

Students did not consider themselves as "at risk" even if they stated that they were "slow or different learners." They expressed a desire for the opportunity to learn the required material. This is illustrated by the statements following: *"I needed more time in courses like science and math. Instead of having to finish some course in one year or semester, I would like to have more*

time and go slower. I do not think I have been taught some of the things I saw on the TAKS."

Additionally, the students were aware that not every teacher was equipped to teach them, as illustrated by this comment from a senior: "*All teachers write, read, grade, but sometimes some teachers do not know how or cannot help kids who are 'slow' or 'different'.* But other teachers explain things in ways everyone can learn. We need teachers who know how to teach kids like us." This view is shared by the Center on Educational Policy (2002), which stated that most states have not taken steps to ensure equitable instruction across the board.

Students in this focus group indicated that some teachers did not have skills necessary to address the different learning styles in their classrooms. Additionally, students wanted more time to learn materials for courses they considered very difficult, like chemistry or physics, and more time to learn concepts that they had not had the opportunity to learn. In the **What If** loop, students revealed their desire and hope for more effective ways of instruction for all students to acquire basic skills—such as offering a plausible alternate plan to the current education system of an extended day or semester for students who are "low achievers" so that they may work at a slower pace to meet curriculum standards and objectives necessary to master the state assessment.

Group I – Summary

Research Question # 1

What critical perspectives do teachers (and students) have about the effects of high-stakes testing?

The theory of perception or the conceptual map expressed by the students in Group I (students who had failed the TAAS/TAKS exam) presented the State

of Texas Assessment System as a flawed process with negative consequences. They perceived themselves as students who were “slow” or different learners, locked within a system that had not always given them the opportunity to learn required materials. Furthermore, they perceived that the system was incapable or unwilling to meet their educational needs, while holding them responsible for state-mandated assessments, and ultimately denying high school diplomas to students who failed these assessments.

Students were firm in their conviction that no single assessment should determine whether a student received a high school diploma. They wanted their coursework to have a greater value in decisions to grant or withhold diplomas.

Research Question #2

How do intended and unintended consequences of high-stakes testing affect students?

Students who failed the state graduation assessments were denied a high school diploma. This fact motivated students to try to pass the state graduation assessment, an intended consequence of high-stakes testing. However, highly negative emotions, such as fear, stress, frustration, and hopelessness may be considered unintended consequences of the state assessment process. These negative emotions impacted students’ ability to learn and their performance on the state assessment.

Next, practice tests were seen as vehicles teachers use to prepare students for the state assessment. Students found they were becoming better test-takers. Students perceived that some teachers were not prepared to teach students who were “different learners.” Students who studied and worked hard to pass the state assessment but failed the TAKS exams were often discouraged,

disappointed, and disillusioned, and they dropped out of school or merely left school after completing the state-required courses without acquiring a high school diploma.

Research Question #3

How do intended and unintended consequences of high-stakes testing affect instruction and classroom practices?

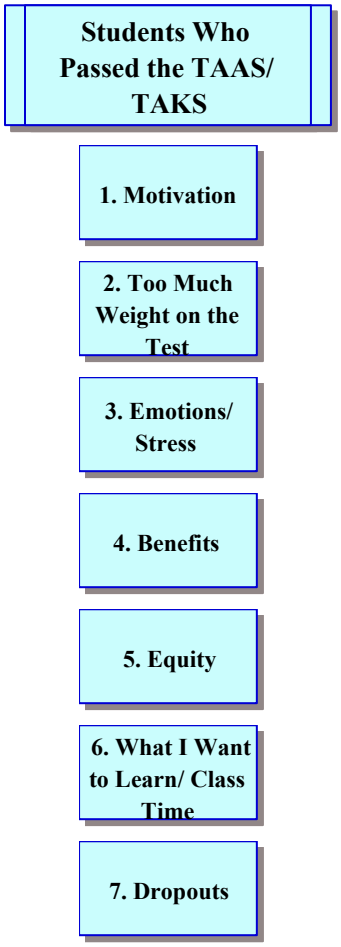
The students revealed that the system sometimes tested students on information that they had not been given the opportunity to learn. The students expressed a need for additional time to master some core concepts and for teachers who were equipped to teach students who were "low achievers," suggesting the school should offer extended semesters for certain academic core courses. These students expressed a desire to eliminate unproductive classroom practices, such as review of disassociated concepts or information. The affinity *Practice Tests,* revealed that students perceive practice tests as a means of obtaining enough information to pass the TAKS. It should be noted that, in this mind map, students did not ask to be exempt from concepts required by the state they had not mastered; instead, they requested additional course time and support to meet these standards, indicating that they were of the expectation that they could and should learn required core assessment objectives. And some students questioned if practice tests were the best way to accomplish this objective.

Group II – Analysis Process

The students in Group II, who had successfully passed the state assessments for graduation, were asked the same question as those students in

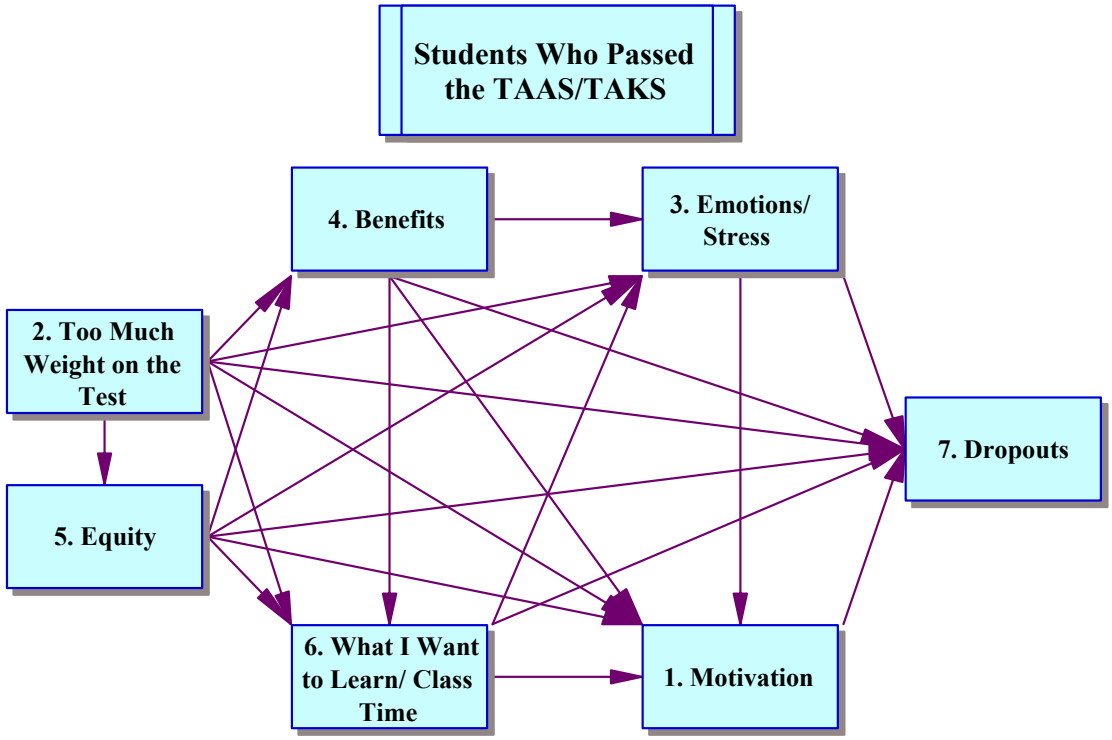
Group I: “Tell me about your experiences with Texas State Assessment Exams, commonly called the TAAS/TAKS.” They produced these affinities:

Figure 5.8
Group II – Affinity Choices



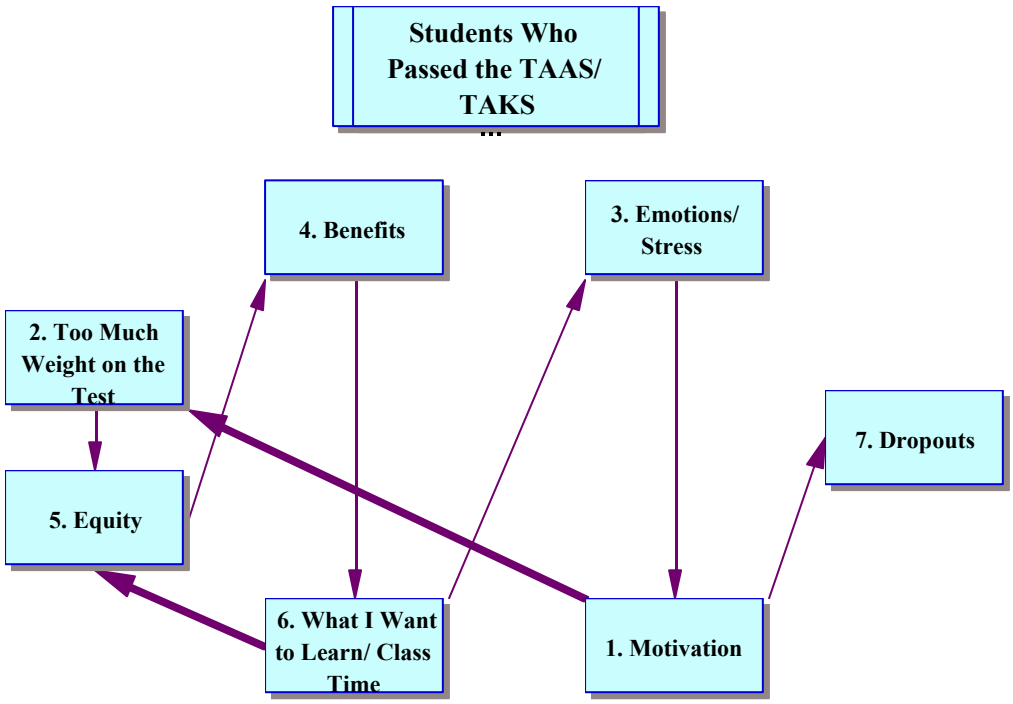
The cluttered SID for Group II containing each link in the IRD is shown below:

Figure 5.9
Group II – Cluttered SID



As with group I, all redundant links are removed according to their delta and SID assignments. The Pareto Protocol was examined for conflicting relationships to ascertain if possible feedback loops exists. The Pareto Chart for Group II follows.

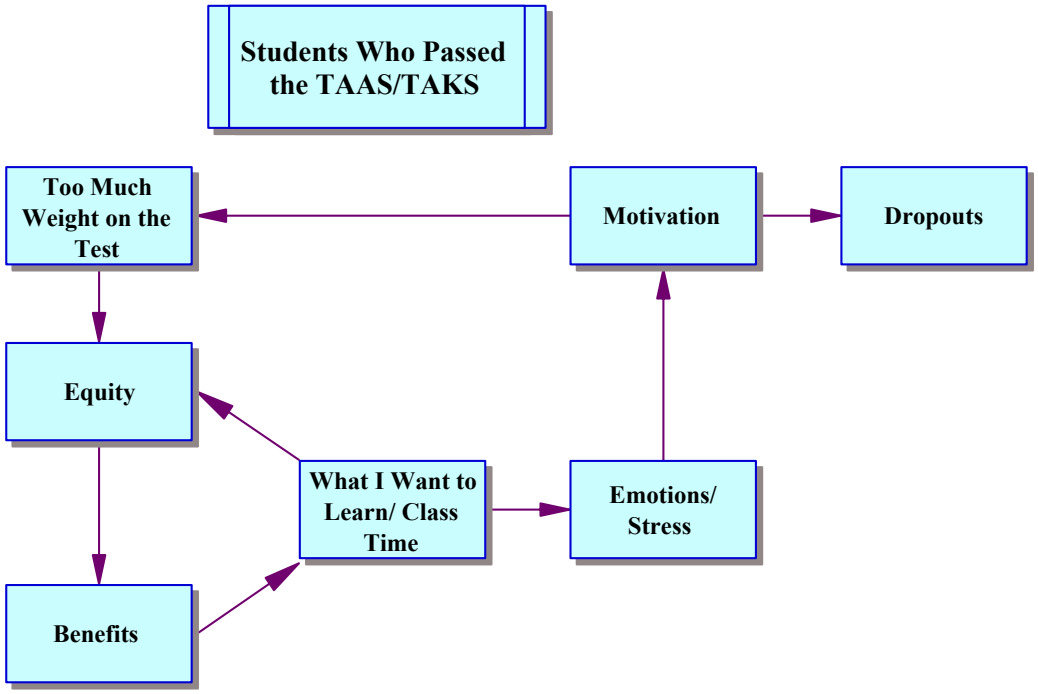
Figure 5.10
Group II – Pareto Reconciled SID



The Pareto chart indicates that there is a feedback loop between the affinities: *Equity, What I Want to Learn*, and *Benefits*.

The conceptual or mind map below represents the theoretical perceptions of the students. A tour of this system revealed that the Texas state assessment was again the main driver or cause of the system.

Figure 5.11
Group II – Theoretical Perceptions

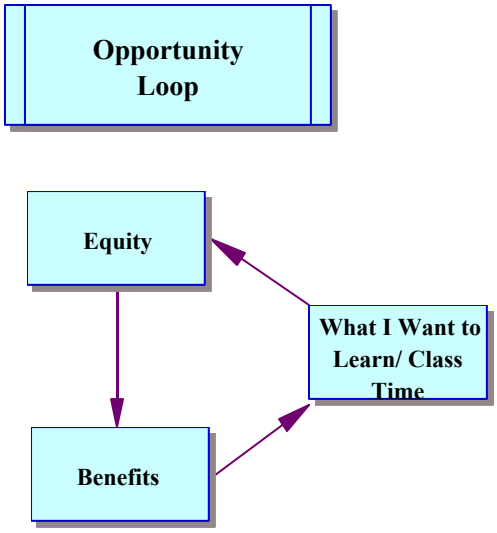


Although all students in this focus group had successfully passed the State of Texas Assessment, they were also quick to point out that one assessment should not determine whether or not a student receives a high school diploma. From their viewpoint, required courses should not be disregarded when determining if a student received a high school diploma. Students stated that course grades were better indicators of what they had learned or accomplished during class. But they agreed the state needed some sort of standard to measure students.

There is a relationship between the affinities *Equity* and *Benefits*, determining the values students attached to their high school learning

experience. The systems flows into a feedback loop, **Opportunity,** which included the *Equity, Benefits*, and *What I Want to Learn.*

Figure 5.12
Group II – Opportunity Feedback Loop



Here, students assessed the Texas Assessment and Accountability System. The Opportunity Loop symbolizes gained and missed opportunities students perceived they had experienced because of the Texas state assessments. The students perceived that the Texas state assessment causes teachers to teach more “to the test” (TAAS), narrowing the curriculum and limiting areas of study students wish to explore and learn. Their view is similar to the assessment made by Meyer, who argued that in high-stakes accountability systems, teachers and administrators are likely to exploit all avenues to improve measured performance. For example, teachers may “teach to the test” (Meyer, 1996, p. 140)

As top students in their school, the Group II students wanted to be sure that they were given every opportunity to learn concepts above the norm, allowing them to be more competitive in college. Denial of such an opportunity was considered to be inequitable. These students stated that they were often bored while reviewing information for the Texas state assessment. And one student remarked, "I*f passing the state assessment is what determines who gets a high school diploma, then why not let kids like us take the state assessment in 9th or 10th grade and go on to college where we could study more things we are interested in?"* Students in this focus group demonstrated a different kind of maturity and wanted to jump through necessary "hurdles" to be successful as quickly as possible.

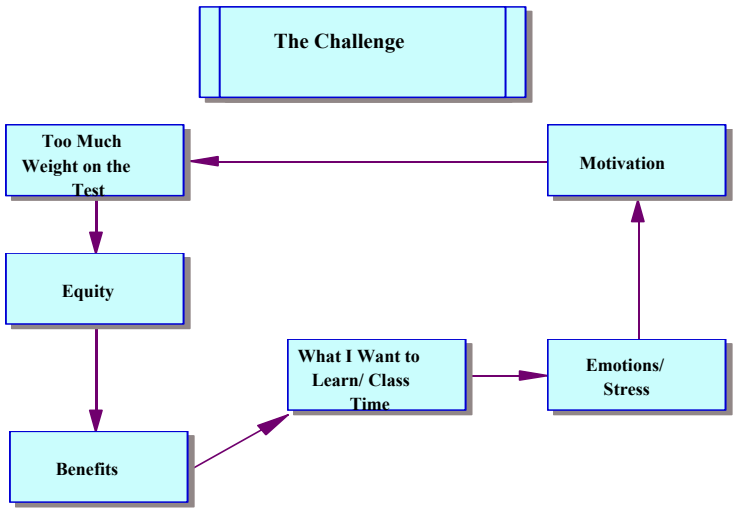
Not being able to study more challenging topics in the curriculum created negative emotions in these students. The students were resentful of having to "mark time" while other students were given time by teachers to meet the state assessment standards. They wanted and expected more from school. Students remarked that some courses were reduced to test materials or isolated facts—resulting in a learning environment that did not meet their needs or promote real learning.

McNeil & Valenzuela (2000) warned that instead of creating an improved learning environment, crude forms of assessment often reduce opportunities to acquire higher learning skills. The students perceived that there must be equity for all: Students who struggle to master the state requirements should be able to move forward at a faster rate if they are able to do so, and be given the opportunity learn more. Furthermore, they believed, all students deserve to be

treated equitably. Equity includes core standards for all students as well as advanced concepts and knowledge.

The affinities *Equity*, *What I Want to Learn* are also apart of a large segment of the system, which included the affinities *Emotions*, *Motivation*, and *Too Much weight on the Test*, called **The Challenge.**

Figure 5.13
Group II – The Challenge Diagram



Students' emotions determined to some extent how motivated they were about course materials. Students found it easy to get interested and involved in concepts they considered challenging. However, they were quick to point out that recurrent, repetitive drills over questions for the state assessments were boring and a waste of class time. Students perceived that benefits from such drills for advanced students were minimal at best.

After students had passed the required state assessments, they were no longer required to come to school on TAAS/TAKS practice test days. Nor did

some students have to participate in class during TAAS/TAKS practice drills. Academically advanced students saw this as a waste of their time and underdevelopment of their abilities. However, students stated that TAKS assessments had made school better in other respects. Teachers now taught a core curriculum to everyone. And they were interested in all children passing the TAKS exams. However, from these students' point of view, the TAAS/TAKS narrowed the curriculum taught. Additionally, their system or theory of perception revealed that highly negative emotions and little or no motivation may cause academically advanced students to drop out of school.

Group II – Summary

Research Question #1

What critical perspectives do teachers (and students) have about the effects of high-stakes testing?

The theory of perception generated by the students in Group II suggested that state assessments were necessary, noting that these standards had help produce better schools and students by promoting common core curriculum standards, goals and objectives. But, from their viewpoint, no one assessment should determine whether or not a student received a high school diploma.

Research Question #2

How do intended and unintended consequences of high-stakes testing affect students?

After passing the state assessment, some high-achieving students were not motivated to apply themselves to their coursework. Passing the exit exam served as a disincentive for these students because they did not have to work as

hard in their class after successfully passing the exam, an unintended consequence of high-stakes testing.

The Group II students perceived that the state assessment narrowed the curriculum and limited areas of study they wished to explore and learn. From their viewpoint, some courses were reduced to test materials or isolated facts, resulting in a learning environment that did not meet their needs or promote real learning. This created inequitable conditions hampering their opportunity to learn advanced concepts skills. These perceptions caused negative emotions about the value of schooling.

Research Question #3

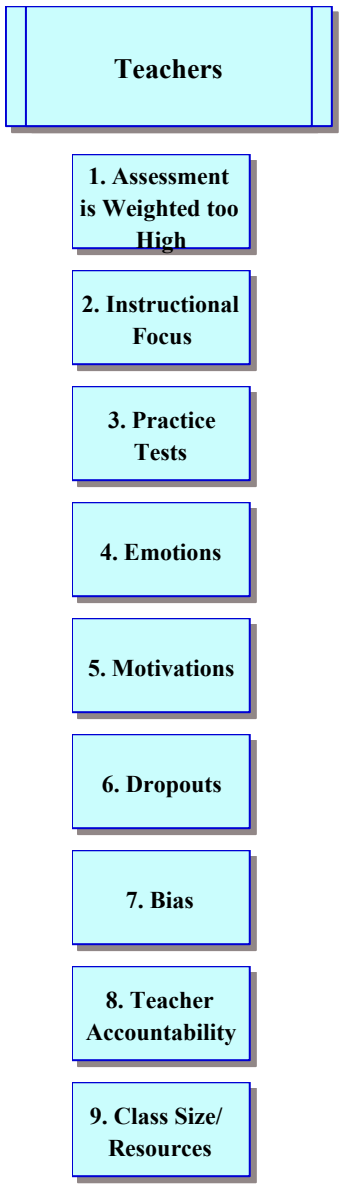
How do intended and unintended consequences of high-stakes testing affect instruction and classroom practices?

The students perceived that assessments dictated what was presented and taught in the classroom, often limiting or eliminating high/advanced level concepts from learning experiences—thus creating a flat, nonproductive, learning environment. From the students' viewpoint, teaching only basic skills was not equitable, nor did it meet the educational needs or the desires of academically gifted students. They viewed many days and hours in school as lost opportunities for "real" learning. Their conceptual map indicates that sometimes they could not see benefits in the current educational programs, which focus on basic skills. The current program produces negative motivation which lowers student morale and desire to learn. Students suggested the educational assessment program be modified to allow testing after each core course, not the yearly assessment of basic skills for those students who had passed the state graduation assessments.

Group III – Analysis Process

Teachers comprised the last group. As noted in Chapter IV, they were asked, “Tell me about your experiences with the Texas State Assessment.” The focus group generated the following affinities.

Figure 5.14
Group III – Affinity Choices



The cluttered SID is produced and the conflicting affinities are reconciled using a Pareto analysis. Finally, the clean SID was produced.

Figure 5.15
Group III – Cluttered SID

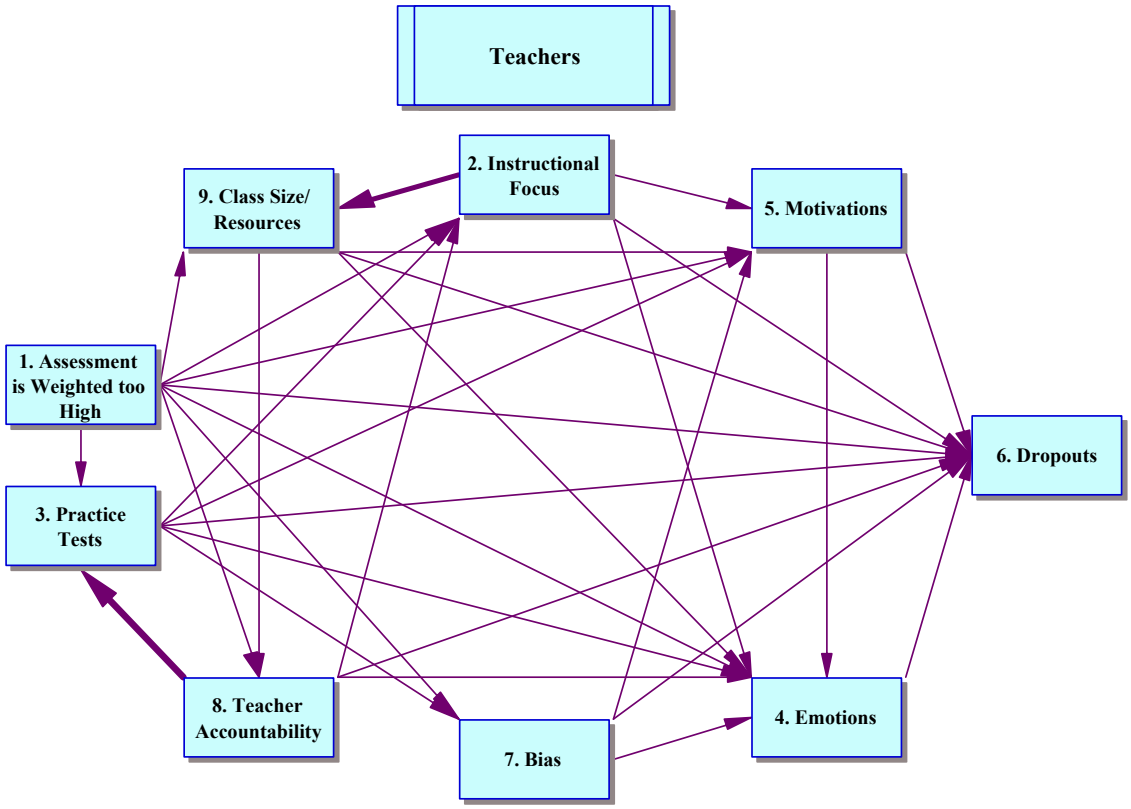
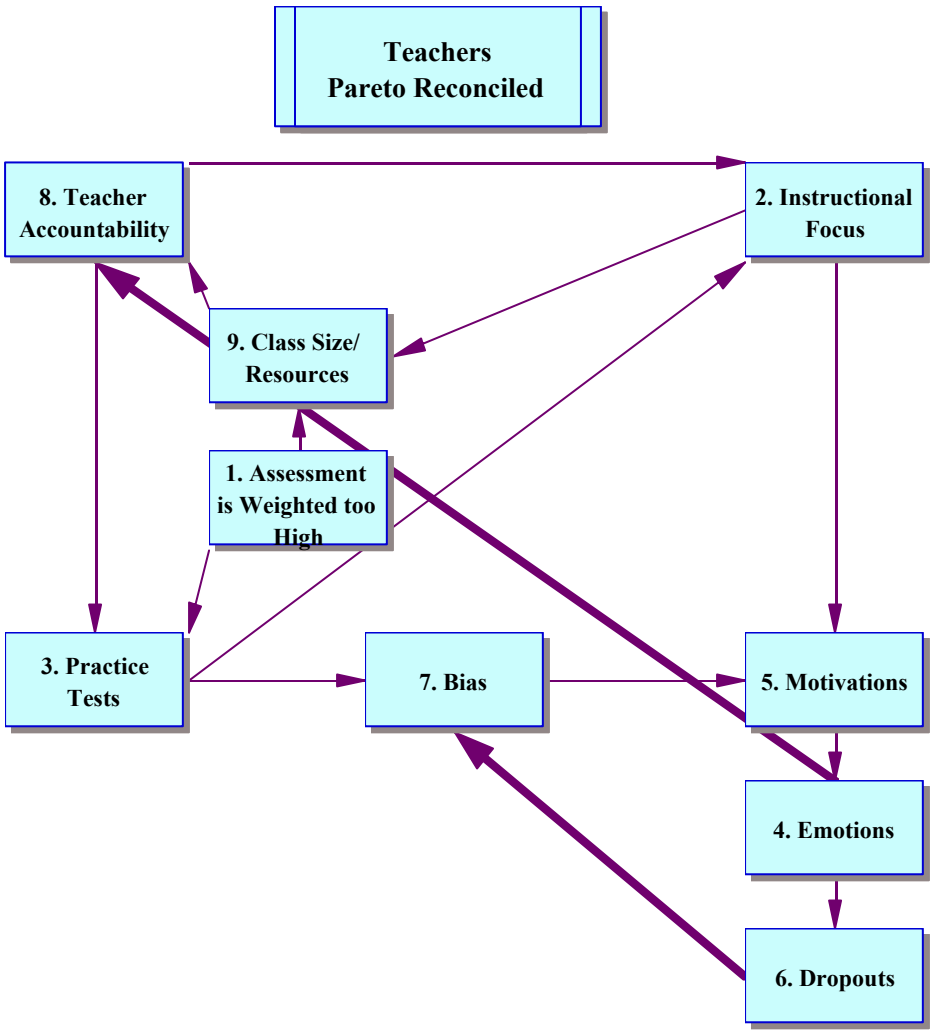
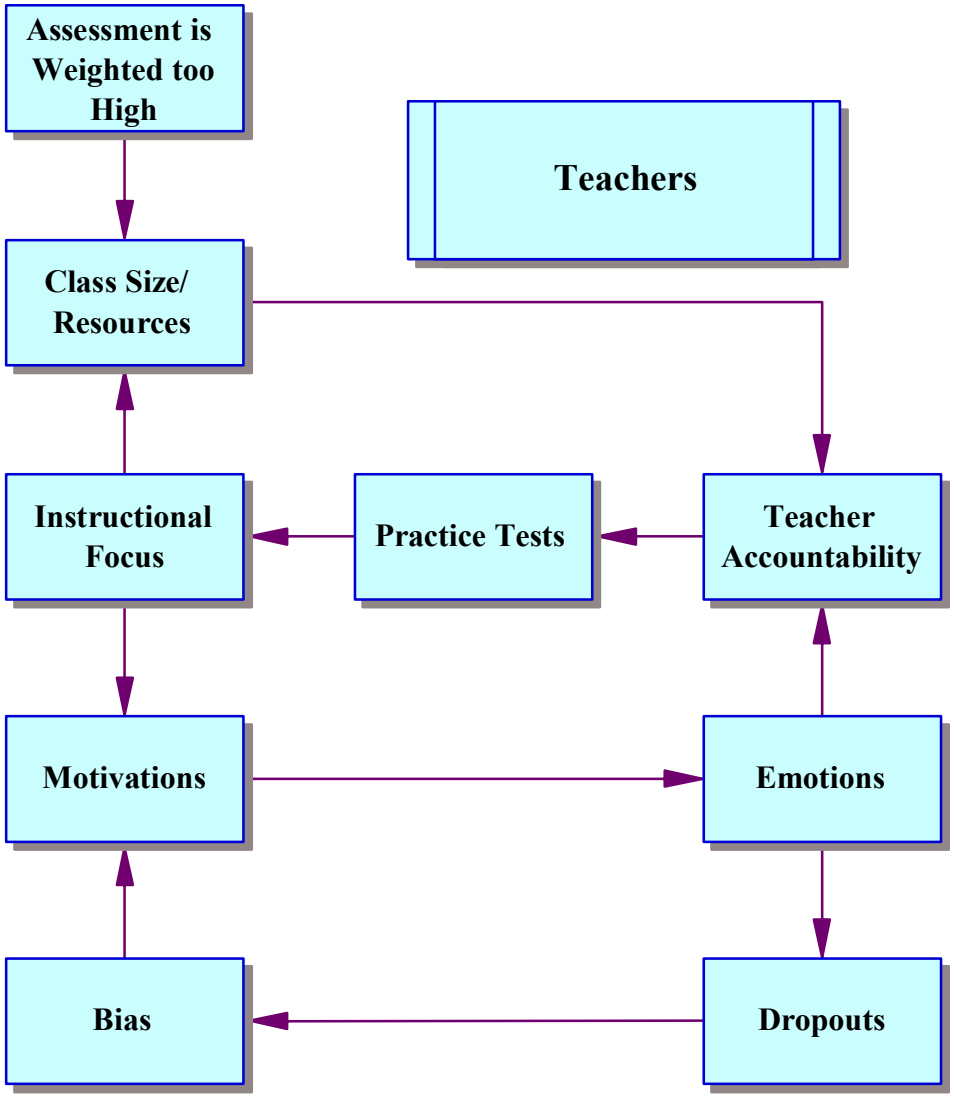


Figure 5.16
Group III – Pareto Reconciled SID



The conceptual or mind map below represents the theoretical perceptions of the teachers. The primary driver of the map is the TAKS assessment being weighted too high. The cluttered SID is produced and the conflicting affinities are reconciled using a Pareto analysis. Finally, the clean SID was produced.

Figure 5.17
Group III – Theoretical Perceptions

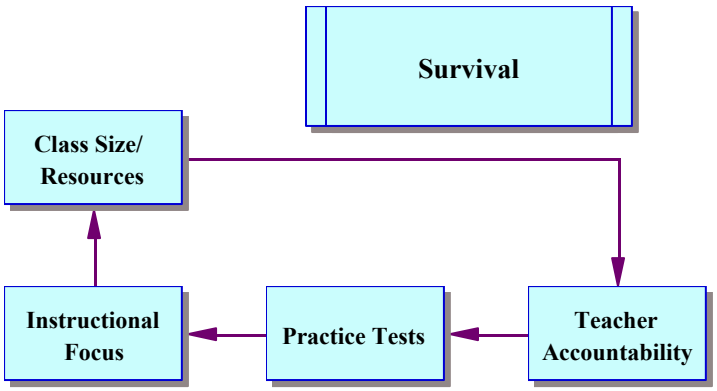


This system contains three loops. Teachers also perceived the Texas state assessments as a primary force or cause. They recognized that the establishment of the TAKS/TAAS assessment was the primary reason why education in the State of Texas had changed, fostering common goals, expectations, and objectives for students and teachers. However, during

interviews, teachers remarked that changes in Texas state education to meet state assessment demands were not always beneficial, sometimes failing to motivate teachers and students to achieve higher academic performance. Some researchers have cautioned teachers and administrators that placing a premium on student test performance had led to instruction that is focused primarily on test preparation, thus limiting the range of educational experiences and reducing the instructional skills of teachers (McNeil, 2000; Smith, 1991).

Class Size and/or Resources begins a feedback loop, called **Survival**, that contains the additional affinities *Teacher Accountability, Practice Tests, Instructional Focus***.**

Figure 5.18
Group III – Survival Feedback Loop



The affinity *Class Size/Resources* affected the affinity *Teacher Accountability*. “To reach kids we need more time, resources, and smaller classes. We need smaller classes for regular and advanced students. It is almost impossible to give each student the type of attention needed in large classes.” The teachers perceived that large classes were causing them not to have enough time to assist certain students, usually the students having the greatest

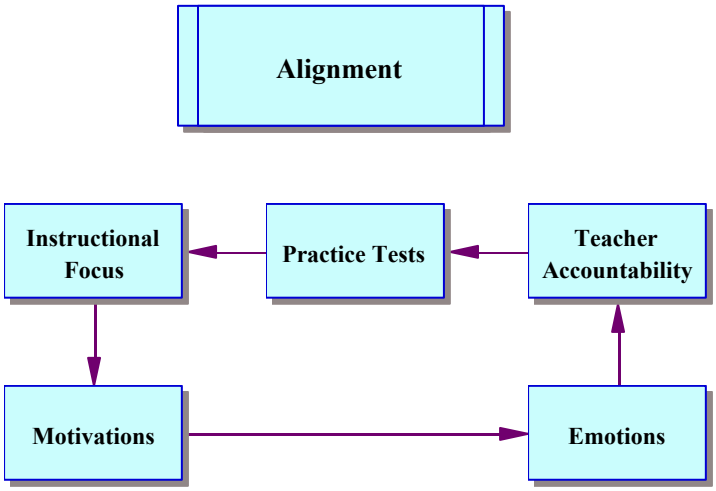
academic difficulty. Some teachers stated they taught to the middle, and teaching the TEKS and curriculum left little time to help those students who needed individualized attention.

The teachers concluded that they needed more resources to support students who were academically challenged. *Teacher Accountability* is the next theme within the loop. Teachers are held to higher teacher accountability by the state, and this accountability has caused teachers to pay more attention to the state curriculum and to teach all types of students—minorities, Special Ed, 504, as well as ESL students. Some teachers were concerned about losing their jobs if their students did not perform well on state assessments. But teachers who were excellent teachers were not concerned about losing their source of employment. Teachers were keenly aware that the Texas state assessment process held them accountable for student test performance, and they perceived that the state assessment was a critical component of classroom instruction.

Next, the mind map demonstrates that teacher accountability influences the practice tests. Practice tests in the system were seen as being positive and negative. Some teachers believed in the old adage “practice makes perfect” and they felt practice tests were required to help students, especially minority, become familiar with the state assessment format and to perform successfully on that assessment. However, most teachers stated that their school held too many practice state assessment exams, averaging 1-2 district practice exams per six weeks. The teachers confirmed that students who had successfully passed the state assessments often felt that these practice tests were a waste of their school time, and wanted the opportunity to learn information that was important to them (advanced topics).

The affinities, *Practice Test*, *Motivation*, and *Instructional Focus* form a loop, called **Alignment.**

Figure 5.19
Group III – Alignment Feedback Loop



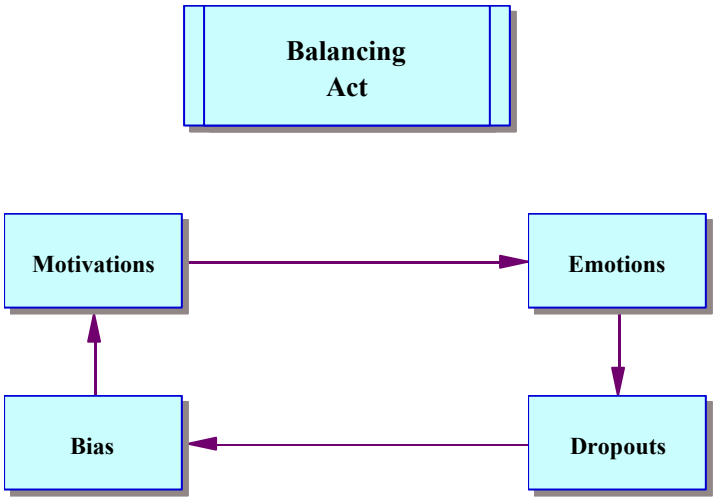
Comments from teachers demonstrate that the affinities *Instructional Focus, Practice Tests* were aligned in this conceptual map to the state assessment. The affinity *Instructional Focus* may produce both positive and negative effects on classroom instruction and practices. Teachers remarked that much of the instruction in some classes was dictated by what was being tested on the Texas state assessments. For many average and below-average classes, it meant "teaching to the test" in an effort to maximize the number of students who would pass the state assessment and maintain or improve school assessment scores. Teachers were quite clear that the state assessments have improved core curriculum and expectations for all students. This is illustrated by the comments of one teacher:

> I am a coach and I grew up playing sports during high school in Texas. It was not uncommon to just pass athletes through school without making

them do school work. Many students did not get an opportunity for an education; the school system simply did not require them to study. As long as the school was winning games, the “good” athletes were never worried about failing classes or not graduating from high school. This unwritten policy was extremely harmful. With the new Texas state assessment, more students are getting a good basic education. All students are expected to pass the state assessment.

The affinity *Motivation* was perceived as a catalyst that could promote learning. It begins the last loop in the system developed by the teachers, which is named **Balancing Act**. This loop includes the affinities *Emotions*, *Dropouts*, and *Bias.*

Figure 5.20
Group III – Balancing Act



The teachers viewed motivation as a student’s willingness to study and perform class assignments. Texas state assessments were instituted to motivate teachers and students to reach higher performance levels. However, teachers in this case study found that the nature of high-stakes testing programs had quite the opposite effect, lowering student morale, and sometime fostering dropouts.

Emotions were viewed as being both positive and negative, affecting both teachers and students. Negative emotions played a significant part in how teachers and students viewed the state assessment. Teachers dealt with pressure from administrators and parents to improve test scores, as well as pressure to limit teaching to what is tested. And some teachers reported changing teaching methods in ways that were not always beneficial to some students or demonstrating the best teaching methods. Both teachers and students often felt stress, anxiety, and/or fear when preparing for and during state assessments. Some students tended to become overwhelmed and drop out or consider dropping out of school. And some teachers tended to move to courses that were not tested by the state. The teachers also admit that some of the advanced students felt little or no anxiety about state assessments, because they were confident they would easily pass the assessments. Positive emotions were experience by teachers when high percentages of their students passed the exams.

Bias was identified as language or cultural bias. Some teachers were aware of cultural or language biases on the exam and/or resources they used in their classrooms. They agreed that minorities, poor, and/or ESL students have missed exam questions because they were not familiar with terms of middle-class America. As one instructor remarked:

> I was never so heart-broken as when I realized that over half of the students in my class missed questions on the TAAS exam because it contains a name or other cultural phases they were not familiar with. Often I feel that I have failed poor African-American or poor Hispanic students when I realize they failed the TAAS exam by one or two questions because the question was biased. Not racially biased but socio-economically biased."

The teachers stated they felt test-makers unintentionally included biased terms and/or phrases on the exam. One teacher remarked that she understood how this was possible because she really did not understand that certain terms were biased until she was analyzing why students who knew certain concepts and performed well on district assessments failed similar questions on state assessments. *"We must eliminate such bias from the state assessment. The stakes are too high, especially when one question may determine whether or not a student receives a high school diploma,"* stated one instructor.

Group III – Summary

Research Question #1

What critical perspectives do teachers (and students) have about the effects of high-stakes testing?

The theory of perception created by the teachers in this case study revealed that consequences of high-stakes testing (assessment) for high school students are both intentional and unintentional. The theory of perception revealed that teachers perceive the Texas state assessment as a change agent of the educational process, with both intentional and unintentional consequences. The potential problem was not with the state assessment, per se, but the instances when the assessment had unintended and potentially negative consequences for individual students, groups of students, or the educational system. A conceptual map of the teachers' responses suggested that instructional focus within the classroom and services offered to students were determined by the schools' or communities' need to be successful on the state assessments. The teachers perceived that the state assessments and demands to teach the concepts on these assessments have narrowed the curriculum and created an undesirable

culture of “teaching to the test” in an effort to improve or maintain their schools’ accountability ratings. The teachers noted that there are benefits from a statewide assessment system, such as core objectives, equal expectations, and education of all students with basic skills, including the poor and minorities. However, state assessments seem to present other negative consequences, including teaching to the test, lowering students’ expectations, elimination of higher-level concepts, or not meeting the educational needs of students, especially low-achiever and gifted students. From the teachers’ viewpoint, no single assessment should determine whether or not a student should receive a high school diploma.

Research Question #2

How do intended and unintended consequences of high-stakes testing affect students?

The state assessment seems to have opposite effects to what it was developed for, in that it does not always motivate teachers and students to higher learning—only higher test-taking skills and abilities. And sometimes the assessment seems to cause certain students to miss learning opportunities and/or drop out of school, thus failing to fully develop the nation’s most valuable asset—the intellectual abilities of its youth.

Research Question #3

How do intended and unintended consequences of high-stakes testing affect instruction and classroom practices?

The teachers perceived that some classroom practices and instructional changes in the classroom are beneficial, and others are not. According to the responses, both students and teachers have benefited from state standards,

goals, and objectives for courses, and they viewed these as positive intended consequences of the state assessment. However, teachers also perceived a narrowing of the curriculum; classes that were too large to adequately provide all students with the support needed for optimum academic development; reduction of time available for ordinary instruction; and loss of control over what was taught, when it was taught and how it was taught, leading to de-skilling and de-professionalism.

System Comparisons and Discussion

When comparing and considering the affinities and their descriptions from each group in the study, similarities and differences were apparent.

Comparative Analysis of Research Question #1

What critical perspectives do teachers (and students) have about the effects of high-stakes testing?

The perception that the TAAS/TAKS assessment was weighted too heavily was the main driving force for each group in the study. Both teachers and students contended that no single assessment should determine if a student should receive a high school diploma. Many researchers oppose using exit exams as a sole criterion for graduation. Professional testing standards recognize that tests are not perfect measures; for the American Education Research Association, for example, states that, “In educational settings, a decision or characterization that will have a major impact on a student should not be made on the basis of a single test score. Other relevant information should be taken into account if it will enhance the overall validity of the decision” (American Educational Research Association et al.., 1999, Standard 13.7, p. 146). The

groups preferred a combination of state assessment and coursework, and voiced their opinion that the Texas assessment for graduation was weighted too heavily. Students in Group I suggested an alternative assessment portfolio. All groups were certain that the state assessment had improved the quality of instruction in classrooms, offering basic skills to all students. Ultimately, the focus groups perceived that any single test should not be the sole determinant of a young adolescent's academic future, reducing the student's life chances for higher education.

Comparative Analysis of Research Question #2

How do intended and unintended consequences of high-stakes testing affect students?

The affinity *Opportunity to Learn* is particularly important. Students in Group I perceived that they had not been given the opportunity to learn required materials assessed on the state exam. Providing students with an opportunity to learn material being tested is at the heart of the debate over high-stakes graduation exams. It is an issue of basic fairness not to punish students for failing to learn materials they have not been given the opportunity to learn. The opportunity to learn does not begin in high school, but it is a shared responsibility among grades K-12. Some students are not able to resolve educational deficiencies, and pass the state assessment without substantial remedial support services.

Hence remediation is an essential part of the opportunity to learn. It is necessary to provide students who have failed the graduation assessment the first time around with instructional support. Professional development for teachers is a necessary element of any meaningful remediation program.

Providing students with more of the same instruction that did not work the first time is not the answer. Teachers must be trained in effective ways to help low-achieving students succeed.

As discussed in Chapter II, court rulings on the constitutionally of high school exit exams—such as *Debra P. v Turlington* (1981, 1984) and *G.I. Forum et al.. v Texas Education Agency* (2000)—have identified specific criteria that educational agencies should use to determine whether high school students have had sufficient opportunity to learn material tested. These criteria include ensuring that students: (1) have been given adequate advance notice of graduation testing requirements; (2) have multiple opportunities to take the test; (3) have been taught the tested skills; and (4) have been provided with an opportunity for successful remediation. Students in Group I perceived that they had not been taught the tested skills and/or given the time to learn the requested information. According to Chudowsky (2002), these criteria—including providing students with opportunity for successful remediation—are often the most difficult for states to achieve and demonstrate in practice.

In the discussion of the affinity *Practice Tests* for Group I (students who had failed the state assessment), and Group III (teachers), the state assessment tests were viewed as having both positive and negative consequences. Thus, they tended to create both intentional and unintentional outcomes. On one hand, practice tests enabled students to become familiar with the state assessment format. The use of large numbers of multiple-choice practice tests had enabled students to become familiar with the state testing format; also, it encouraged students to memorize facts, sometimes isolated facts, and had improved the schools' accountability score. However, practice tests eroded hours of class

times and were not seen as a beneficial tool for some students, especially the academically advanced students.

The affinity *Motivation* emerged as another key issue for the three focus groups. How to motivate students is a major concern for teachers, especially in regard to those students who have successfully passed the Texas state assessments and want to explore more advanced concepts, when the instructional focus in many classes was devoted to achieving successful student achievement on the TAAS/TAKS exams. Much of the motivation for high school students to achieve academic success is grounded in extrinsic motivation, namely grades; however, learning is intrinsically motivating (Thompson & Thornton, 2002, p. 787). What must occur is a redirection of motivation so that an intrinsic value of learning is realized. As Zemke and Zemke point out, ". . . learners are more motivated when thy see relevance to their learning, when they have ownership in the process, and when they feel it is tailored to meet their immediate interests and needs" (Zemke & Zemke, 1988, p 59). It is the initial motivation of internal personal issues that must be addressed in order to nurture the external personal part of one's life. Thus students must see the Texas state assessment as a relevant part of their learning, and the schools must be able to address the needs of advanced students, if their learning is to be promoted.

The affinity *Dropouts* was identified as an outcome for all groups. Students and teachers viewed dropouts as a by-product, an unintentional consequence of state assessments. Dropouts were triggers for several reasons according to the groups' mind maps. These included emotions; class time and instructional focus; support and/or motivation. There is also a spirited debate among researchers about whether graduation testing causes increased dropout rates. On one hand,

it appears that many low-achievers start to disengage from school well before graduation tests loom. On the other hand, there are reputable scholars who argue credibly that fear of failing a graduation test increases the likelihood that low-achievers will leave school (Clarke *et al.*., 2000). Also, the current climate of accountability places new pressures on schools to increase student pass rates, which in turn can lead to increased and/or understated dropout rates (Shrag, 2000). Unfortunately, this critical issue is complicated by a lack of uniformity among the states in defining and counting dropouts (Viadero, 2000). However, the mind maps produced in this case study suggest that the current state assessment process may increase the number of dropouts.

The affinity *Emotion*, influences each group's attitude toward learning. Emotions may be positive or negative. The majority of emotions experienced by the groups were negative, including fear, frustration, stress, hopelessness, and high anxiety. A small amount of anxiety is to be expected when students prepare and/or take exams. However, when students are overwhelmed by negative emotions and find it difficult to apply themselves, can impede the learning process.

Comparative Analysis of Research Question #3

The affinities *Use of Class Time, Instructional Focus* and *What I want to Learn* and *More Time to Learn Required Materials* overlapped. All groups in the study viewed what would be a better utilization of class time. But they had different views of how class time should be used. Students in Group I (students who had failed the state assessment) wanted more time to master required materials assessed on the state exam, as well as an opportunity to learn information they perceived they had not been taught. Students in Group II (who

had passed the state assessment) perceived that the state assessment dictated what was presented and taught in the classroom, often limiting or eliminating high/advanced-level concepts from learning experiences. From the students' viewpoints, teaching of only basic skills was not equitable, nor did it meet the educational needs or the desires of academically gifted students. The teachers in Group III perceived a narrowing of the curriculum. This narrowing of the curriculum, along with the confusion of training to pass a test with broader notions of learning and education, can have especially problematic side effects in high-stakes testing for low-income students. The poor, more than their advantaged peers, need not only the skills that training provides, but need the more important benefits of learning and education that allow for full economic and social integration in our society (Amrein & Berliner, 2002). All groups wanted a system to meet their diverse learning needs.

CHAPTER VI – CONCLUSIONS AND RECOMMENDATIONS

Conclusions and Broad Implications

High-stakes testing has been promised to America as a mechanism to foster educational reform, and as an efficient way of assessing school and student performance. However, as this study has found, it can also be a barrier to graduation and/or educational advancement.

In the literature review, it was noted that America has been involved in education for 150 years. After 100 years, with the court case of *Brown v Topeka* in 1956, African-Americans and Hispanics obtained meaningful inclusion in the education process. During the next fifty years, minorities have been given various forms of universal access to education. The Leave No Child Behind Act (NCLB) of 2002 intended to begin a new era of universal proficiency.

Many states are adopting exit exams in an effort to make high school diplomas "have value," namely, to validate that the holder has the knowledge and skills needed to do well in a job or college. State policymakers are adopting these exams to respond to a public outcry that the quality of education has slipped, and that too many students are graduating without standard, adequate skills (Grissmer & Flanagan, 1998).

Based on the input in this study from both the students and teachers, one can conclude that, if high-stakes testing does not promote learning, it must be reevaluated and changed. The purpose of attending high school is to master significant knowledge. Custodial care clearly should not be a part of the effort. Otherwise, America will continue to under-educate its children.

The use of tests in decisions about students' graduation is intended to serve educational policy goals, such as setting high standards for students, closing the achievement gap, raising student academic levels, ensuring equal educational opportunity, fostering parental involvement in student learning, and increasing public support for schools—thus supposedly strengthening and improving our nation's educational system. However, that goal would seem to be flawed when students have not been given the opportunity to learn the information tested.

Determining whether high-stakes testing of students produces better overall educational outcomes requires that its potential benefits be weighed against its potential unintended negative outcomes. This study has revealed that some teachers and students perceive that standards-based reform and high-stakes testing have exceptionally strong impact on African-Americans, Hispanic, and poor students.

Issues that emerged with the current assessment policies when applied in the context of the study involved both unintended and intended consequences. The negative or unintended consequences for students included:

- Use of test scores as the sole criterion for educational decisions, such as graduation
- Narrowed curriculum
- Diminished post-high school educational opportunities for students
- Reduced opportunities to master higher learning skills
- Reduction of class time for ordinary instruction

- Instructional focus in the classroom being determined by what questions are perceived to be on the state assessment, not important skills and knowledge
- State assessment failed to accurately measure student achievement
- Adequate opportunity and time to learn information on the state assessment
- Possible increase in school dropouts

As a result, some students are being denied high school diplomas, and many schools' failure to expose these students to the knowledge and skills students need to pass the Texas graduation assessment test is being highlighted. In this respect, it is particularly interesting that the most significant common bond of the students in the study was poverty, and its negative effects on their education and performance during high-stakes testing. Amrein and Berliner warn that if the unintended consequences of high-stakes assessments are not addressed and corrected they will continue to foster a sub-class or under-class of undereducated students along racial, ethnic, and class lines (Amrein & Berliner, 2002).

It is important to note that students and teachers in this study also perceived high-stakes testing as the catalyst of many positive educational reform measures such as state-wide standards and goals, improved instruction, improved test scores for students, and motivating schools to try to better educate African American, Hispanic, and poor students. However, the theories of perception produced by students or teachers did not indicate that students in the study perceived that high-stakes testing increased learning. Students complained

of learning isolated facts in their preparation for the state assessment. As McNeil and Valenzuela in their Chapter II maintain, “pressure to raise TAAS scores leads teachers to substitute commercial TAAS-prep materials for substance of the curriculum . . . Subjects tested by TAAS (reading, writing, and mathematics) are reduced, in the test and test-prep materials, to isolated facts and fragments of facts. This artificial treatment of these isolated components may enable children to recognize the component on a multiple choice test, but not necessarily enable them to use the components in contexts” (McNeil & Valenzuela, 2000).

Advocates of state assessment who insist that the assessment will invariably motivate students to work harder are not true students of human nature. The premise of state assessment incorporates the underlying notion that students who study and work hard will be successful. This position is correct only if all students begin in equitable resources and in equal environments. Some African-American, Hispanic, and poor students who have labored, studied, and mastered high school course requirements continue to fail the state assessment. Students who are discouraged, disappointed and disillusioned due to high-stakes testing policies are more likely to become depressed and drop out of school rather than be motivated to study harder. We must use scaffolding learning to ensure that students are given the opportunity to learn and master required information tested during the assessment, in a positive learning environment. But, more importantly, policymakers, parents, and other stakeholders must realize that the opportunity to learn does not begin in high school; it begins in kindergarten. Thus when students fail the state assessment, it is not just a student failure, nor simply the failure of the high school teacher; it is the failure of

our educational system, for we clearly have not met the needs of these students. And we are sending out in our community someone who is likely to be dependent rather than independent.

Blanket criticism of high-stakes testing does not appear to be justified. Indeed, progress has been made. High-stakes assessments motivate teachers to make an effort to teach all students, and foster the development of common curriculum, state standards, and goals.

However, to deny that high-states testing does result in withholding diplomas for students who have not had the opportunity to learn the required materials and that current assessment polices do cause irreparable harm to these children and their communities is to commit a serious error. The current assessment system requires a mid-course correction that allows all students the opportunity to learn standards and to provide them the required resources for implementation of this goal. This will not be an easy task for Texas. Texas educational reform is rooted in two distinct conflicts: unequal distribution of resources among Texas school districts, and the challenge to the hegemony of the state's traditional agricultural and oil interests (Carnoy, 2000).

Instead of creating an improved learning environment and promoting learning, graduation assessments may reduce students' opportunities to obtain higher learning skills, particularly in the case of low-income students (McNeil & Valenzuela, 2000). The findings of this study imply that curriculum standards, instruction, and state assessments are not aligned. Texas and other states continue to make progress in this endeavor. However, to ensure fairness, the impact of state assessments must be minimized until the assessment system can be amended to be equitable for all. Additional research is needed to understand

the impact of high-stakes testing on students, and specifically African-Americans, Hispanics and the poverty-level students.

From a culturalist standpoint, the knowledge that every child benefits from high expectations and standards is not a problem for the African-American and Hispanic community. However, the implementation of high-stakes testing is an issue. The failure to account for race, culture, ethnicity, and economics in the discourse of educational policy in reference to high-stakes policies fails to provide open opportunities for discourse and debate of factors and issues considering high-stakes testing, its implementation, and its impact on African-American, Hispanic, and poor students.

Ultimately we must answer the question whether any single test should be the sole determinant of a young adolescent's academic future. The standards movement and high-stakes testing present both opportunities and risks to minorities and poor students. These students are among those who stand to benefit most if all students receive high-quality instruction and are doomed to a lifetime of failure if they do not. In other words, testing is perceived to be an efficient way of assessing school and student performance, but it can also be a barrier to graduation and educational advancement.

Legislators and policymakers have tended to sanction the use of high-stakes testing with the intent to use this assessment process to set high standards for student learning, raise student achievement-levels, and ensure equal educational opportunity (Amrein & Berlin, 2002; Carnoy & Smith, 2001). However, this case study has demonstrated that for some students and teachers these goals are not being realized. They have experienced unintended consequences of high-stakes testing including teachers teaching to the test,

lowering the standards of student learning, fostering dropouts, and the underdevelopment of students' academic ability. The process has apparently been producing students with high test-taking skills but under-developed problem-solving and critical-thinking skills.

Moreover, quantitative and qualitative research studies examining the validity of the above statements and expectations in support of high-stakes testing have found exactly the opposite effects (McNeil & Valenzuela, 2000; Orfield & Krohaber, 2001; Sacks, 1999; Sheldon & Biddle, 1998). NCLB reflects the consensus that all children can learn high standards if they are given the means to do so. The law rejects what President Bush has called the "soft bigotry of low expectation" (Measuring what matters: An update on educational assessment and accountability, 2002) and demands that all children be given the opportunity to learn.

Schools must be staffed with qualified teachers who can and are willing to teach low-achievers. Teachers must be given the time to teach required curriculum objectives and students must have adequate time to learn the information. Districts must be given the funding required to provide equal opportunities for all students to learn, including meaningful and timely remediation, professional development, and counseling. Public schools should partner with colleges and universities and use research to promote and improve student learning. Legislators must do more than have a vision and mandate change; they must provide the adequate resources to fund educational reform. As Joel Baker (Bowsher, 2001) reminds us, "A vision without action is merely a dream. Action without a vision just passes time. Vision with action can change the world" (p. 259). Graduation assessments are not standards. We can and

must change assessment policies such that they promote “real” learning and “real” progress. If unintended consequences of high-stakes assessments are not addressed and corrected, they will continue to foster a sub-class or under-class of undereducated students along racial, ethnic, and class lines. We must educate all children to standards that will allow them to lead independent, successful lives.

Specific Strategic Recommendations

The nation’s economic strength and social cohesion depend on all children being well-educated. Improving the achievement of African-American and Hispanic students will help to eliminate economic disparities and ensure that all young people are well-prepared to become active, productive citizens.

Many states are adopting high school exit exams as a part of a larger strategy to raise student achievement in public schools. Policymakers have assured the American public that high-stakes testing will increase student learning. However, high-stakes testing can and evidently does cause negative unintended effects, such as encouraging certain ineffective instructional practices and using assessments to withhold diplomas.

When state assessments have high stakes for individuals, policymakers have an equally high level of responsibility to ensure that assessments are used fairly, that their content is not culturally biased, and that their effects are not discriminatory. A student performance may vary depending upon a variety of factors. Test scores are not ends in themselves, but a means of measuring how well students are learning knowledge and skills that are the real ends of education. The following strategies are recommended for states and schools using high-stakes testing and which wish to promote “real” learning.

- **Well-Defined Assessments**. The state must make sure that assessments are valid, fair, reliable, free of language and/or cultural bias. Assessments must be aligned with standards and curriculum. The state testing system should be revised to include a variety of item formats that measure both basic and more advanced knowledge skills. Test items should reward good teaching.
- **Graduation Exams.** Consideration should be given to the elimination of graduation exams as a sole assessment to determine who is granted or denied a diploma. The consequences of high-stakes testing for individual students are often posed as an either-or proposition. Denial of a high school diploma alone is not an effective treatment for low achievement. Schools can use strategies to reduce the need for this either-or choice, such as coupling early identification of such students with effective remediation.
- **Improved and Challenging Curriculum Standards.** Efforts should be made to ensure that curriculum standards are challenging, worthwhile, coherent, and focused starting in elementary schools. Efforts should be made to strengthen school policies, counseling, and academic support to encourage and not discourage African-American and Hispanic students from taking rigorous academic courses beginning in elementary and middle school.
- **Instruction**. Instruction must be a positive reflection of the state's and school's intellectual compass within a framework of common values and goals representing intellectual progress and

achievement that look beyond high school graduation. It must provide clear expectations and rigorous, coherent, culturally relevant learning experiences that promote higher-level thinking skills and problem analysis with parallel remediation for those students in need. Instruction must allow students to explore and make meaning of phenomena by researching and/or weighing evidence, expressing diverse points of view, and seeing how things are connected or related in the context of their lives and environment. Finally, there must be continuity between curriculum, instruction, and assessment, which fosters classroom interactions, community involvement, student work productions, and measurable student achievement and learning.

- **Well-Qualified Teachers.** Schools should be provided with well-qualified teachers. Efforts should be made to strengthen preparation, induction, and professional development program for teachers. Additionally, salaries for teachers should be increased such that teachers are paid on a par with other professionals.
- **Teacher Incentives.** Incentives should be provided for highly qualified teachers to teach in high-poverty and high-minority schools.
- **Redirect Teachers.** Teachers should be cautioned not to teach to the test, but to teach to the standard.
- **Professional Development.** Teachers should be provided with professional development in effective ways to help students master

the content in the state assessment process and to address the diverse learning styles of students within their classrooms.

- **Reduce Class Size.** Sustained reductions should be made in class sizes in high minority and/or low-income schools
- **Extended Learning Opportunities.** Students should be provided with extra time and the attention they may need to master the academic knowledge and skills contained in state standards. Policymakers should acknowledge that some students may not have been given the opportunity to learn vital concepts in elementary or middle school necessary to pass state graduation assessments. Recognizing this must include remediation and intensive instruction during the school day, and/or alternate programs such as extended semester or after-school programs.
- **Instructional Support.** Teachers should be provided with various kinds of support to help them improve classroom teaching and student learning.
- **Accurate Assessment of English Language Learners.** English Language Learners are particular vulnerable to potential negative consequences when high-stakes decisions are based on test. Testing procedures should be modified in a manner to provide appropriate accommodation for the effect of limited English proficiency on the subject matter being tested, while maintaining the validity and comparability of test results among all students.

- **Expand Preschool.** Efforts should be made to provide universal access to high-quality state-funded preschool programs, including readiness instruction for minority students.
- **Dropout Counseling Center.** Student centers should be established to identify probable dropouts and provide programs to help students remain in schools.
- **University and Public School Partnerships.** Support additional research and case studies to learn more about the impact of high-stakes and learning.
- **Resource Disparities.** Resource disparities must be eliminated. Districts and schools should be provided with the funds necessary to implement educational reform.

Additional Research

"It was the best of times, it was the worst of times"— this quotation from Charles Dickens' A Tale of Two Cites describes the position many African-Americans, Hispanics, and poor children experience during this era of high-stakes testing and educational reform. A single state graduation assessment in many states determines who is granted or denied a high school diploma. More minorities and poverty-level students are in quality schools than ever in the history of America, achievement gaps are narrowing, and high-stakes testing mandates are ensuring that teachers and schools are focused on the education of all children. However, the number of students who are failing state graduation assessments and who are being denied a high school diploma after struggling for 12 or more years within an educational system is unacceptable. Many minority students and their parents perceive that they are not being given the opportunity

to learn required materials. Districts, schools, teachers often do not have adequate resources needed for instruction and remediation. Policymakers perceive education does not meet past expectations and it is too expensive.

Additional research is required to understand the unintended and intended consequences of high-stakes testing. Millions of dollars and billions of man-hours have been invested in high-stakes assessment. And we do not fully understand the effects of high-stakes state assessment policies on the lives of children, nor on the intellectual wealth of our nation. Additional research should encompass scrutiny and evaluation of such aspects as the following:

- Practices, policies, and implementation involved in high-stakes testing that use a sole criterion to determine promotion or graduation
- Implementation of assessment policies
- Alignment of curriculum, instruction, and assessment
- Large-scale assessment instruction and supporting materials
- Elements of student learning and assessments: cognition, observation, and interpretation.
- Opportunities to learn and master state assessment objectives
- Remediation
- Educational funding and assessment
- Cost analysis

Regardless of what research is performed, we must answer three important questions: (1) Are high-stakes testing policies increasing and promoting learning? (2) Have all students been given the opportunity to learn, including classroom resources, adequate funding, quality instruction, and parallel

remediation? (3) What financial impact do high-stakes graduation policies have on our economy? We have the knowledge to improve the education of America's children. We can and should find alternatives to failure. We must enable all children to be successful.

These recommendations are not just for African-American or Hispanic students; to the contrary, they will benefit all children in the pursuit of universal proficiency and educational excellence. Equal opportunity and educational excellence are sometimes cast as competing priorities, but we should approach them as complementary parts of a unified approach to educational reform. The reward to America will be long-term economic and social benefits for the entire nation.

APPENDIX A

Group Warm-Up

All high school students in Texas are required to the take the Texas State Graduation Exit Assessment. We need your assistance in understanding the impact of these assessments on you and your classroom activities. In a few minutes, I am going to ask you to tell me about your experiences with Texas State Assessment Exams, commonly called TAAS/TAKS.

- To begin, find a chair and get as comfortable as you can.
- Close your eyes.
- Put aside your thoughts of the day.
- Take a deep cleansing breath.
- Now imagine yourself in class preparing for the state assessment.
- Taking TAAS/TAKS practice tests.
- Taking the TAAS/TAKS
- Think about your surroundings while preparing or taking the test (long pause).
- Allow yourself to become aware with all of your senses.
- Focus on what it feels like to be totally absorbed while preparing or taking the TAAS/TAKS exam.

Now, tell me about your experiences with the state assessments (**Issue Statement).** Reflect on all the thoughts you have concerning state assessments. Write these thoughts down on the cards. Please write one thought per card. All thoughts are OK. Don't analyze; just write. No one will criticize your thinking.

APPENDIX B

Student Focus Group I
Interview Protocol
Respondent
Axial Coding

Thank you for participating in this interview. Group I has identified several common themes or affinities that describe their experience with the TAAS/TAKS exams. Let us look at each affinity once more while you tell me about your experiences with each.

1. All or Nothing

The group described ***All or Nothing*** as the State of Texas Graduation Policy. To receive a high school diploma, students must pass the state's assessment exams. Tell me about your experiences with the Texas graduation exam.

2. **Emotions**

Emotions represented how the group felt about preparing and taking the State of Texas Assessments for high school graduation. Tell me how you felt when preparing or taking the TAAS/TAKS.

3. **Dropouts**

Dropouts identified students who have left school before graduation due to academic or state assessment failure or students who complete required courses for high school graduation but fail the state assessment. ***Tell me about this component***.

4. **More time to learn**

This affinity reflects students' desire for and time to master core State of Texas requirements. Tell me about *M**ore time to learn***.

5. Practice Test

This affinity included district-wide TAAS/TAKS practice tests; pre-tests for the district TAAS/TAKS practice test; and benchmark testing. Tell me about your experiences with ***Practice tests***.

6. **Material tested**

Students identified material tested as specific concepts examined on the test and when or if these concepts were taught during their middle and high school career. Tell me about your experiences with ***Material tested***.

7. **Use of Class time**

This affinity referred to the instructional focus of day-to-day classroom instruction and activities. Tell me about how you spent time in class in reference to the TAAS/TAKS test.

APPENDIX C

Student Focus Group II
Interview Protocol
Respondent
Axial Coding

The group has identified several common themes or affinities that describe their experience with the TAAS/TAKS exams. Let us look at each affinity once more, while you tell me about your experiences with each.

1. Motivation

The students described this affinity as stimulus or desire to learn course materials and prepare for state assessments. Tell me about this affinity.

2. Too much weight on the test

The group described ***Too much weight on the test*** as the State of Texas graduation policy. To receive a high school diploma, students must pass the state's assessment exams without considering coursework students have completed. Tell me about your experiences with this policy.

3. Emotions

Emotions represented how the group felt about preparing and taking the State of Texas assessments for high school graduation. Tell me how you felt when preparing or taking the TAAS/TAKS.

4. Benefits

This affinity reflects perceived beneficial outcomes of the Texas State Assessments. Tell me about perceived benefits for the state assessment.

5. **Equity**

This affinity included district and state policies that address the educational needs of students. Tell me about your experiences with equity.

6. **Dropouts**

Dropouts identified students who have left school before graduation due to academic or state assessment failure or students who complete required courses for high school graduation but fail the state assessment. Tell me about this component.

7. **What I want to learn/Use of class time**

This affinity refers to the instructional focus of day-to-day classroom instruction and activities. Tell me about how you spent time in class in reference to the TAAS/TAKS test.

APPENDIX D

Teachers Focus Group III
Interview Protocol
Respondent
Axial Coding

The group has identified several common themes or affinities that describe their experience with the TAAS/TAKS exams. Let us look at each affinity once more, while you tell me about your experiences with each.

1. TAAS/TAKS Assessment Graduation Value

The group described ***All or Nothing*** as the State of Texas graduation policy. To receive a high school diploma, students must pass the state's assessment exams and the true measure of the assessment. Tell me about your experiences with this affinity.

2. **Instructional Focus**

The affinity ***Instructional focus*** describes instructional practices and curriculum decisions based on the school's need to receive a successful rating from the Texas accountability system and how these changes affect student learning. Tell me about this affinity.

3. **Emotions**

Emotions represented how the group felt about preparing students to take the State of Texas Assessments for high school graduation. Tell me about emotions.

4. **Practice Tests**

This affinity included district-wide TAAS/TAKS practice tests; pre-tests for the district TAAS/TAKS practice test; benchmark testing; and regular tests for courses. Tell me about your experience with this affinity.

5. **Motivation**

The teachers described this affinity as a stimulus or desire, or a student process they created within students to learn course materials and prepare for state assessments. Tell me about motivation.

6. Dropouts

Dropouts identified students who have failed the TAAS/TAKs assessment and have left school before graduation. Tell me about this affinity.

7. Bias

This affinity reflects cultural and language barriers inherent in instruction, class materials, and Texas graduation assessment. Tell me about your experience with this affinity.

8. Teacher Accountability

This affinity represents State of Texas standards of excellence for all teachers, and the impact the state assessment has on teacher accountability. Tell me about this affinity.

9. Class Size/Resources

This affinity represents resources provided for the educational development of students, including class size. Tell me about your experience with this affinity.

APPENDIX E

Combined Interview Affinity Relationship Tables and Conflicting Relationships for Groups I through III

Table E.1
Group I – Combined Interview
Affinity Relationship Table
(Students Who Failed the TAKS)

Affinity Pair Relationship		Affinity Pair Relationship
1 → 2		4 ← 5
1 → 3		4 ← 6
1 → 4		4 → 7
1 → 5		5 <> 6
1 → 6		5 → 7
1 → 7		6 → 7
2 → 3		
2 <> 4		
2 ← 5		
2 ← 6		
2 <> 7		
3 ← 4		
3 ← 5		
3 ← 6		
3 ← 7		

Table E.2 Group I – Conflicting Relationships				
Affinity Pair Relationship		**Affinity Pair Relationship**		**Affinity Pair Relationship**
1 ← 4		2 → 5		4 → 5
1 ← 5		3 → 6		4 → 6
1 ← 6		4 ← 7		

Table E.3 Group II – Combined Interview Affinity Relationship Table (Students Who Have Passed the TAAS/TAKS)		
Affinity Pair Relationship		**Affinity Pair Relationship**
1 ← 2		4 ← 5
1 ← 3		4 → 6
1 ← 4		4 → 7
1 ← 5		5 → 6
1 ← 6		5 → 7
1 → 7		6 → 7
2 → 3		
2 → 4		
2 → 5		
2 → 6		
2 → 7		
3 ← 4		
3 ← 5		
3 ← 6		
3 → 7		

Table E.4 Group II – Conflicting Relationships				
Affinity Pair Relationship		**Affinity Pair Relationship**		**Affinity Pair Relationship**
1 → 2		3 → 6		5 ← 6
1 → 5		4 ← 6		
2 ← 5		4 → 5		

Table E.5 Group III – Combined Interview Affinity Relationship Table (High School Teachers)				
Affinity Pair Relationship		**Affinity Pair Relationship**		**Affinity Pair Relationship**
1 → 2		3 → 4		6 ← 7
1 → 3		3 → 5		6 ← 8
1 → 4		3 → 6		6 ← 9
1 → 5		3 → 7		7 <> 8
1 → 6		3 ← 8		7 <> 9
1 → 7		3 <> 9		8 ← 9
1 → 8		4 ← 5		
1 → 9		4 → 6		
2 ← 3		4 ← 7		
2 → 4		4 ← 8		
2 → 5		4 ← 9		
2 → 6		5 → 6		
2 <> 7		5 ← 7		
2 ← 8		5 <> 8		
2 → 9		5 ← 9		

Table E.6 Group III – Conflicting Relationships

Affinity Pair Relationship		Affinity Pair Relationship		Affinity Pair Relationship
3 < 7		4 > 8		6 > 7
4 > 5		4 > 9		8 > 9
4 > 7		5 > 7		

REFERENCES

AERA position statement concerning high-stakes testing in pre-K-12 education. (2002). Retrieved October 2, 2002, from http://www.aera.net.about/policy/stakes.htm

Amrein, A., & Berliner, D. (2002). *The impact of high-stakes tests on student academic performance: an analysis of NAEP results in states with high-stakes tests and ACT, SAT, and AP test results in states with high school graduation exams*. Retrieved December, 2002, from http://edpolicylab.org

Amrein, A., & Berliner, D. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives,* 10(18), 1-70.

Appropriate use of high-stakes testing in our nation's schools. (2003). Retrieved August 22, 2003, from http://www.apa.org/pubinfo/testing.html

Bolman, L. & Deal, T. (1997). *Reframing organizations: Artistry, choice, and leadership*. San Francisco: Jossey Bass.

Bowsher, J. (2001). *Fix schools first: Blueprint for achieving learning standards*. San Diego: Aspen Publishers.

Brookhart v Illinois Board of Education (1983).

Cannell, J. (1987). *Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average*: Daniels, WV: Friends of Education.

Carnoy, M. (2001). *School vouchers: What does the evidence show?* Washington, DC: Economic Policy Institute.

Carnoy, M., & Smith, T. (2001). *Do higher state test scores in Texas make for a better school outcome*. University: Consortium for Policy Research in Education.

Carnoy, M., Loeb, S., & Smith, T. (2000). *Do higher state test scores in Texas make for better high school outcomes?* Paper presented at the American Educational Research Association Annual Meeting. New Orleans: Consortium for Policy Research in Education.

Cattarall, J. (1989). *School dropouts: Here today, here tomorrow.* Los Angeles: UCLA Center of Research and Evaluation, Standards and Student Testing.

Chief State School Administrators. *Designing accountability systems: Toward a framework and process*. (2002). Council of Washington: Chief State School Administrators.

Chudowsky, N., Kober, N., Gayler, K. & Hamilton, M. (2002). *State high school exit exams: A baseline report*: Center on Education Policy.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Weinfield, A. et al.. (1966). *Equality of educational opportunity*. Washington: U. S. Printing.

Corbett, H., & Wilson, B. (1990). *Unintended and unwelcome: The local impact of state testing.* Boston: American Educational Research Association.

Corbett, H., & Wilson, B. (1991). *Testing, reform, and rebellion*. Norwood, N.J.: Ablex.

Cremin, L. (1989). *Popular education and its discontents*. New York: Harper & Row.

Darling-Hammond, L. (1988). Accountability and teacher professionalism. *American Educator, 12*, 8-13.

Debra P. v Turlington, 474 F Supp 244 (M D Fla. 1979).

Delgado-Gaitian, C. (1993). Researching change and changing the researcher. *Harvard Educational Review, 63*(4), 389-411.

Digest of education statistics 2000. (2001). National Educational Statistics US Department of Education. *Digest of education statistics 2000*. Washington: National Academy Press.

Dropouts keeping students in schools. (2002). Retrieved October 10, 2002, from www.law.harvard.edu/civilrights

Elazars, D. (1984). *American federalism: A view from the states*. New York: Harper & Row, Publisher.

Ellwein, M., Glass, G. & Smith, M. (1988). Standards of competence: Proposition of the nature of testing and reform. *Educational Researcher, 17*, 4-9.

Elmore, R., Abelmann, C, & Fuhrman, S. (1996). The new accountability in state education reform: From process to performance. In H. F. Ladd (Ed), *Holding schools accountable: Performance-based reform in education* (pp. 65-98). Washington: The Brookings Institution.

Equity driven achievement: Focus school districts. (2000). Austin: Charles A . Dana Center. University of Texas.

Fetterman, D. & Smith, C. (2002). *Improving student performance: Partnering for success.* Paper presented at the National Conference Education Accountability.

Forgione, P. (1999) Achievement in the United States: Are students performing better? Hearing for the Committee on Education and the Workforce. Washington D.C.

Foster, C. (1985). *Teaching and learning in elementary schools*. New York: MacMillan Publishing Company.

Freire, P. (1970). *Pedagogy of the oppressed*. New York: Continuum.

Gall, J., Gall, M., & Borg, W. (1999). *Applying educational research: A practical guide*. New York: Longman.

GI Forum v Texas Education Agency (87 F. Supp. 2d 667 2000).

Greenwald, R., Hedges, L., & Laine, R. (1996). Interpreting research on school research and student achievement. Review of Education Research 66(3), 411-416.

Griffin, B. Heridon., M. (1996). An examination of the relationship between minimum competency test performance and dropping out of high school. *Education Policy Analysis Archives, 18*(3), 243-252.

Grissmer, D., & Flanagan, A. (1998). *Exploring rapid achievement gains in North Carolina and Texas*. Washington: National Education Goals Panel.

Grissmer, D., Flanagan, A., Kawata, J, & Williamson, S. (2000). *Improving student achievement: What NAEP tests tell us*. Retrieved October 10, 2002, from http://www.rand.org/publications/MR/MR929

Haney, W. (1999). *Report for testimony in GI forum v Texas education agency*. Boston: Boston College, School of Education.

Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives, 8*(41).

Haney, W. (2001, January 13, 2001). *Revisiting the myth of the Texas miracle in education: Lessons about dropout research and dropout prevention.* Paper presented at the Dropout Research: Accurate Counts and positive Interventions, Cambridge, MA.

Hanushek, E. (1989). The impact of differential expenditures on school performance. *Educational Research, 18*(4), 45-65.

Hanushek, E. (1994). *Making schools work: Improving performance and controlling cost*. Washington: The Brookings Institution.

Hanushek, E. (1996). A more complete picture of school resource policies. *Review of Educational Research, 66*(3), 397-410.

Hauser, R. (1993). Trends in college entry among Whites, Blacks, and Hispanics. In C. R. Clotfelter (Ed.), *Studies of supply and demand in higher education* (pp. 61-104). Chicago: University of Chicago Press.

Hauser, R. (1999). "Should we end social promotion? Truth or consequences." In G. K. Orfield, M. (Ed.), *Raising standards or raising barriers? Inequality and high-stakes testing in education*. New York: The Century Fund.

Hedges, L., & Greenwald, R. (1996). Interpreting research on school resources and student achievement: A rejoinder to Hanushek. *Review of Educational Research, 66*(3), 411-416.

Hedges, L., & Greenwald, R. (1992). Does money matter: Meta-analysis of studies of the effects of differential school input on student outcomes. *Educational Research, 23*(3), 5-14.

Heubert, J., & Hauser, R. (1999). *High-stakes: Testing for tracking and promotion, and graduation,* from http://www.nap.edu/html/highstakes

Hillard, A. (1976). *Alternatives to I. Q. testing: An approach to the identification of gifted minority students*: California State Department of Education.

Hot topics: Assessment. (2002). *Education Week*.

Hurwitz, N. & Hurwitz., S. (2000). Do high-stakes assessment really improve learning? *American School Board Journal, 87*(1), 21-25.

The impact of high-stakes testing policies on minority and disadvantaged students. (2000). *American Youth Policy Forum*.

Jacob, B. (2001). Getting tough? The impact of high school graduation exams. *Education Evaluation and Policy Analysis, 23*(2), 99-121.

Jacobson, J. (2002). High stakes testing : Is it worth the risks, from http://www.smar\terkids.com/recenter/library/article.asp?article=1295&redir=true

Klein, S., Hamilton, L., McCaffery, D. & Stecher, B. (2000). *What do test scores in Texas tell us?* Santa Monica: RAND.

Koretz, D. (1988). Arriving at Lake Wobegon: Are standardized tests exaggerating achievement and distorting instruction? *American Educator, 12*(2), 46-52.

Kreitzer, A. Madaus., G. & Haney, W. (1989). Competency testing and dropout. In L. Weis, Farrar, E. & Petrie, H. (Ed.), *Dropout from schools: Issues, Dilemmas and Solutions*. Albany: State University of New York Press.

Kuykendall, C. (1992). *From rage to hope: Strategies for reclaiming Blacks and Hispanic students*. Bloomington, IN: National Education Services.

Ladd, H. (1996). *Holding schools accountable*. Washington: The Brookings Institution.

Ladson-Billings, G. (1994). *The dreamkeepers*. San Francisco: Jossey-Bass.

Ladson-Billings, G., & Tate, W. (1997). Toward a critical race theory of education. *Teachers College Record, 97*(1).

Linn, R. (2000). Assessment and accountability. *Education Research, 29*(2), 4-16.

Linn, R., Grause, M., & Sanders, N. (1990). Comparing state and district results to national norms: The validity of the claims that "everyone" is above average. *Educational Measurement: Issues and Practice, 9*(3), 5-14.

Locke, L., Spirduso, W., & Silverman, S. (1993). *Proposal that works. A guide for planning dissertations and grant proposals*. Newbury Park, CA: Sage.

Madaus, G. (1988). The influence of testing on curriculum. In L. Tanner (Ed.), *Critical issues in curriculum: 87th yearbook of the NSSE part 1*. Chicago: University of Chicago Press.

Madaus, G., & O'Dwyer, L. (1999). A short history assessment: Lessons learned. *Phi Delta Kappan, 80*(9), 688-695.

Making standards matter. (1999). *American Federation of Teachers*.

Marshall, C., & Rossman, G. (1994). *Designing qualitative research*. Thousand Oaks, CA: Sage Publication.

McLaughlin, M., & Shepard, L. (1995). *Improving education through standards-based reform: A report of the National Academy of Education*. Stanford: National Academy of Education.

McNeil, L., & Valenzuela, A. (2000). The harmful impact of the TAAS system of testing in Texas: Beyond the accountability rhetoric. In G. K. Orfield, M. (Ed.), *Raising the standards or raising barriers? Inequality and high-stakes testing in public education*. New York: The Century Fund.

Measuring what matters: An update on educational assessment and accountability. (2002). New York: Committee for Economic Development.

Mertens, D. (1998). *Research methods in education and psychology: Interacting diversity with quantitative and qualitative approaches*. Thousand Oaks: Sage.

Meyer, R. (1996). Comments on chapters two, three, and four. In H. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 137-145). Washington: The Brookings Institution.

Montecel, M., Supik, J., & Cardenas, J. (1994, October 1). Improving student performance: Study identifies better approach*. IDRA Newsletter*.

Murnane, R., & Levy, F. (1996). *Teaching the new basic skills*. New York: The Free Press.

A nation still a risk: An education manifesto (1998) from http://edexcellence.net/library/manifes.html

Natrellio, G. & Pallas, A. (1999). *The development impact of high-stakes testing. Paper presented at the conference on Civil Rights implications of high-stakes testing.* Paper presented at the Harvard Civil Rights Project, Teacher College and Columbia Law School.

Neill, M. & Gayler, K. (1999, December 4, 1998). *Do high-stakes graduation tests improve learning outcomes? Using state-level NAEP data to evaluate the effects of mandatory graduation tests.* Paper presented at the High-stakes Testing Conference, Columbia University.

Northcutt, N. & McCoy, D. (2002). *Book chapters on interactive qualitative analysis*: The University of Texas at Austin.

Ohanian, S. (1999). *One size fits few: The folly of education standards*. Portsmouth, NH: Heineman.

Olson, L. (2002, January 9, 2002). Testing system in most states not ESEA ready. *Education Week*.

Orfield, G. Krohaber, M. (2001). *Raising standards or raising barriers? Inequality and high-stakes testing in public education*. New York: The Century Foundation Press.

Peer reviewer guidance for evaluating evidence of final assessments under Title I of the Elementary and Secondary Education Act. (1999). United States Department of Education.

Popham, W. (1987). The Merits of measurement-driven instruction. *Phi Delta Kappan, 68*(9), 679-682.

Popham, W. (1999). Why standardized tests? *Educational Leadership, 56*(6), 8-15.

Quality counts. (2001). from http://www.ed.week.org/sreport/qc01

Rodgers, N., Paredes, V, & Mangino, E. (1991). High-stakes minimum skills tests: Is their use increasing achievement? Paper presented at annual meeting of the American Education Research Association: ERIC Document Reproduction Service (No. ED 336 422).

Sacks, P. (1999). *Standardized minds: The price of America's testing culture and what we can do to change it.* Cambridge, MA: Perseus Books.

Scheurich, J. (2002). *The history of critical race theory*. Retrieved December 1, 2002, from http://www.edb.utexas.edu/faculty/scheurich/proj7/crthistory.htm

Schmoker, J. (2002). The results we want. *Educational Leadership, 57*(5), 62-65.

Sheldon, K., & Biddle, B. (1998). Standards accountability and school reform: Perils and pitfalls. *Educational Measurement, Issues and Practice, 9*(3), 15-33.

Shepard, L. (1990). Inflated test score gains: Is the problem old norms or teaching the test? *Educational Measurement, Issues and Practice, 9*(3), 15-22.

Shepard, L. (1992). *Will national tests improve student learning? (CSE Technical Report 342)*. Los Angeles: University of California, Center for Research on Evaluation Standards and Student Testing.

Shepard, L., & Smith, M. (1989). *Flunking grades: Research and policies on retention in education*. London: Falmer.

Shrag, P. (2000). Too good to be true. *American Prospect, 11*(4).

Simons, H. (1980). *Towards a science of the singular: Essay about case study in educational research and evaluation.* Norwich, UK: University of East Anglia, Centre for Applied Research in Education.

Smith, M., Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement, Issues and Practice, 10*(4), 7-11.

Soder, R. (1996). *Democracy, education and the schools*. San Francisco: Jossey-Bass.

Stake, R. (1995). *The art of the case study research*. Thousand Oaks, CA: Sage.

Stout, R., Tallerico, M., & Scribner, K. (1995). Values: The 'what'? of the politics of education. In K. L. Scribner, Donald (Ed.), *The study of educational politics*. Washington: The Falmer Press.

Sum, A. (1999). *Literacy in the labor force*. Washington: National Center for Education Statistics.

Tatum, B. (1999). *"Why are all the Black kids sitting together in the cafeteria?" and other conversations about race.* New York: Basic Books.

Texas Assessment of Knowledge and Skills (TAKS) - Update. (2002), from www.tea.state.tx.us/student assessment.

Thompson, B. & Thornton, H (2002). The transition from extrinsic to intrinsic motivation in the college classroom: A first year experience. Education, Vol. 122, Issue 4, 785-792.

Understanding dropouts: Statistics, strategies and high-stakes testing. (2001). Washington: National Research Council.

The use of tests when making high-stakes decisions for students: A resource guide for educators and policy makers. (2000). Washington: U.S. Department of Education.

Valencia, R., Valenzulea, A., Sloan, K. & Foley, D. (2001). Let's treat the cause, not the symptoms: A response to "Thinking carefully about equity and accountability" by Scheurich, Skrla, and Johnson. *Phi Delta Kappan, 83*(4), 318-323.

Viadero, D. (2000). Testing systems in Texas yet to get final grade. *Education Week*.

Vinson, K., Gibson, R., & Ross, E. (2001). *High-stakes testing and standardization: The threat to authenticity*. Retrieved June 24, 2003, from http://www.pipeline.com/~rgibson/highstakestesting.htms

What the AERA says about high-stakes testing. Retrieved March 5, 2003, from www.edletter.org/past/issues/2000-so/area.shtml

Widgor, A. & Garner, W. (1982). *Ability testing: Uses, consequences and controversies (Part I: Report of the Committee on Ability Testing)*. Washington: National Research Council.

Yerkes, R. (1921). *Psychological examining in the United States Army . Memories of the National Academy of Sciences XV*. Washington: U.S. Government Printing.

Yin, R. (1984). *Case study research: Design and methods*. Newbury Park, CA: Sage.

Zemke, R. & Zemke, S. (1988) Thirty things we know for sure about adult learning. Training, pp. 57-61

VITA

Helen Louise Johnson was born in Palestine, Arkansas, the daughter of Walker and Ruby Fleming. After completing her work at Lincoln High School in Forrest City, Arkansas, she attended the University of Arkansas at Pine Bluff, completing a degree in Biological Sciences with a minor in Education. Attending the State University of New York at Binghamton, she obtained graduate hours in Computer Science. In September 2000, she enrolled at Prairie View A&M University, Prairie View, Texas, graduating in 2002 with a Master of Arts degree in Educational Leadership.

During her very successful professional career she has worked in education and business. As a teacher and school administrator she developed and implemented several successful programs for teachers and students. She was awarded the National Science Foundation Scholarship for Gifted Teachers.

In her tenure with International Business Machines as a staff programmer, she was awarded IBM's prestigious Quality Excellence Award, IBM Means Service Award, IBM Director's Award, and IBM/ISSC Re-Architecture Team Award. In 1995, Helen returned to her first love, the education of children, working for the Austin Independent School District. Currently, as a District Curriculum Specialist, she works with over 100 schools developing and implementing district, state, and federal programs. She has gained experience and skills in the area of science instruction and education, student assessment, and accountability. And she continues to hone her shills in the area of diversity education and instructional programs for students in at-risk situations. Her commitment to improving school for African-American children and other

culturally diverse students led her, beginning in June 2002, to seek a doctoral degree in Educational Administration from the University of Texas at Austin.

With almost twenty years of experience in program development, science, technology, assessment, and accountability, she has been involved in several projects and endeavors to help all children attain their educational goals. She believes that schools are places of endless positive possibilities for all children and their communities.

Currently, she resides in Austin, Texas with her husband, Andrew Johnson, and family.

Permanent address:	16813 Cree Lake Leander, Texas 78641
Typing:	Creative Services